Biochimica et Biophysica Acta 1834 (2013) 725–738

Contents lists available at SciVerse ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

Intrinsically disordered regions of p53 family are highly diversified in evolution

Bin Xue [a], Celeste J. Brown [b], A. Keith Dunker [c], Vladimir N. Uversky [a,d,e,*]

[a] Department of Molecular Medicine, College of Medicine, University of South Florida, Tampa, FL 33612, USA
[b] Department of Biological Sciences, Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID 83844-3051, USA
[c] Center for Computational Biology and Bioinformatics, Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202, USA
[d] Byrd Alzheimer's Research Institute, College of Medicine, University of South Florida, Tampa, FL 33612, USA
[e] Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia

ARTICLE INFO

Article history:
Received 28 November 2012
Received in revised form 28 December 2012
Accepted 11 January 2013
Available online 22 January 2013

Keywords:
Intrinsically disordered proteins
Protein evolution
Protein–protein interactions
Protein–DNA interaction
p53 family

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.

1. 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 *D*NA *B*inding *D*omain (DBD), and the C-terminal region (Fig. 1A) [2]. The N-terminal region can be further subdivided into *T*rans*A*ctivation *D*omain *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, TFIIH, Mdm2, RPA, CBP/p300 and CSN5/Jab1 [1]. The CTD of p53 interacts with GSK3β, PARP-1, TAF1, TRRAP, hGcn5, TAF, 14-3-3, S100B(ββ) 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

Abbreviations: ASA, solvent accessible surface area; CDF, cumulative distribution function; CH, charge-hydropathy; 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

* Corresponding author at: Department of Molecular Medicine, University of South Florida, 12901 Bruce B. Downs Blvd. MDC07, Tampa, FL 33612, USA. Tel.: +1 317 278 6448; fax: +1 317 278 9217.
E-mail address: vuversky@health.usf.edu (V.N. Uversky).

1570-9639/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.bbapap.2013.01.012

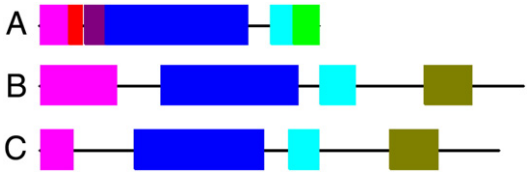

Fig. 1. Domain organization of the members of human family of p53 proteins: p53 (A), p63 (B), and p73 (C). In p53, there is *Trans*Activation *D*omain *1* (TAD1, residues 1–40, pink bar), TAD2 (residues 40–60, red bar), a proline-rich region (PR, residues 64–92, dark pink bar), the central *D*NA *B*ind *D*omain (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).

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 immunoglobin-like β-sandwich that facilitates DNA binding (Fig. 2A and B).

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 *I*ntrinsically *D*isordered *P*roteins (IDPs) and *I*ntrinsically *D*isordered *P*rotein *R*egions (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,

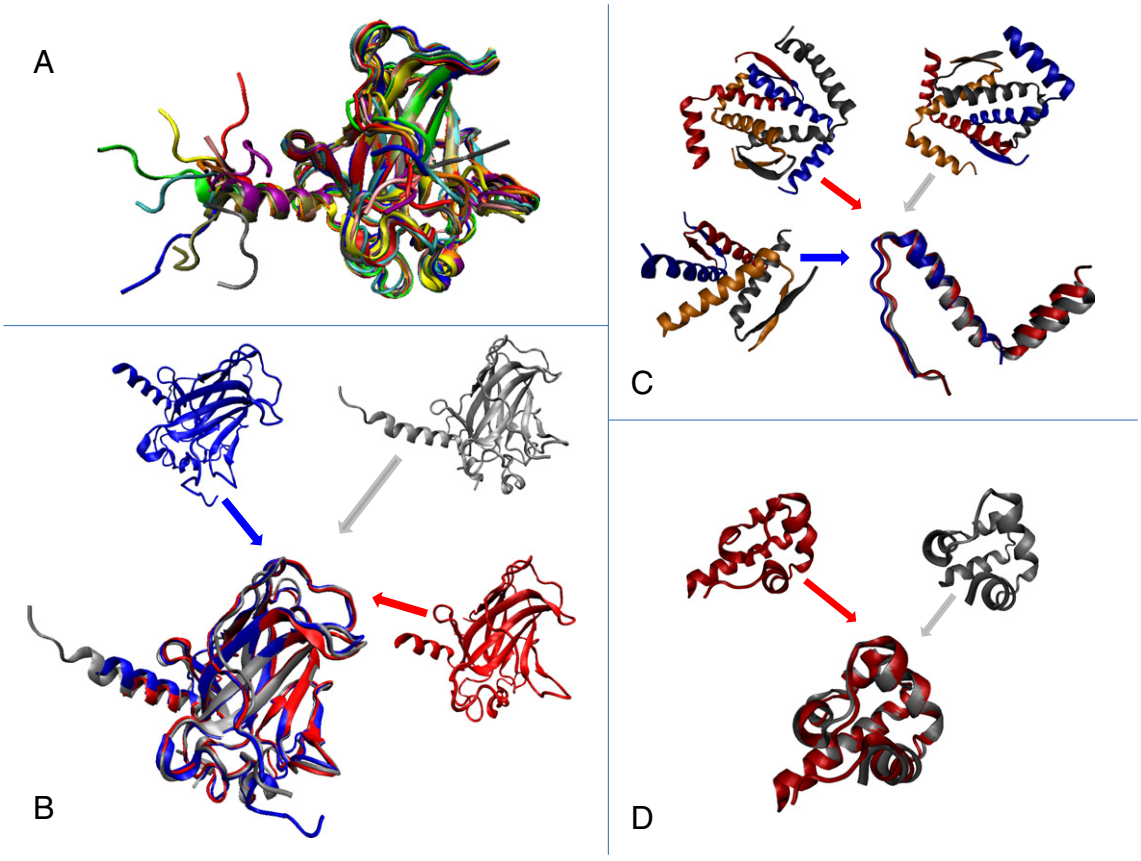

Fig. 2. Structural similarity between the ordered domains of the members of p53 family. A. NMR solution structure of the DBD of p53 (PDB ID: 2FEJ). Representative members of the conformational ensemble are shown as structures of different color. B. 3-D structures of the DBDs of p53 (PDB ID: 2PCX; blue 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: 2WQJ; 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].