Published in final edited form as:

Trends Biochem Sci. 2009 December ; 34(12): 594–600. doi:10.1016/j.tibs.2009.07.007.

Protein–protein interaction networks: how can a hub protein bind so many different partners?

Chung-Jung Tsai¹, Buyong Ma¹, Ruth Nussinov^{1,2}

¹Center for Cancer Research Nanobiology Program, SAIC-Frederick, Inc., NCI-Frederick, Frederick, MD 21702, USA

²Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

Abstract

How can a single hub protein bind so many different partners? Numerous studies have sought differences between hubs and non-hubs to explain what makes a protein a hub and how a shared hub-binding site can be promiscuous, yet at the same time be specific. Here, we suggest that the problem is largely non-existent and resides in the popular representation of protein interaction networks: protein products derived from a single gene, even if different, are clustered in maps into a single node. This leads to the impression that a single protein binds to a very large number of partners. In reality, it does not; rather, protein networks reflect the combination of multiple proteins, each with a distinct conformation.

How can so many interactions be accommodated, given the low estimate of human coding genes?

Networks, gene networks [1], residue networks [2–4], cellular networks [4,5], pharmacological networks [6,7], drug-therapy networks [8], functional networks [9–11] and, in particular, protein–protein interaction networks are the focus of intense research. The number of organisms and systems for which protein–protein interaction networks have been worked out is increasing rapidly. They are displayed in numerous scientific reports and text books depict these well known spider web-like diagrams [12,13]. A node routinely denotes a ‘protein’ and a connecting edge a ‘protein–protein interaction’ or simply ‘interaction’ (Figure 1). In all such interaction networks, most proteins interact with one or only a few other proteins; however, a relatively small number of proteins (hubs) have a very large number of interacting partners, reaching tens and even hundreds. In *Saccharomyces cerevisiae*, among the 4087 different proteins identified with high confidence by mass spectrometry [14], a core set of 7123 protein–protein interactions involving 2708 proteins were identified, with an average of 4.9 subunits per complex, and 5.26 interactions per protein. The per protein distribution of the interactions follows an inverse power law, indicating a typical scale-free network topology [14,15]. Analysis of interacting motifs of hub proteins using databases such as DIP, MIPS, HPRD, BIND along with high-throughput experiments indicated that yeast hubs with multiple interacting motifs are more likely to be essential than those with only one or two [16]. Structural analysis suggested that the more binding sites a protein has, the more essential it is [17]. Alternatively, it was argued that hub

essentiality relates to its number of partners: the more partners, the higher the probability of the hub engaging in essential protein–protein interactions [18]. In principle, the large number of interactions can be accommodated via distinct binding sites; considering the number of domains in the proteins and the combinatorics of the linear domain arrangements, this number can be huge. As each linear domain combination has a different sequence, each constitutes a gene, which implies a very large number of genes. Yet, current data suggest that the number of genes in the human genome is much smaller than the number of all possible domain combinations [19]. Although the exact number of human protein-coding genes in the genome remains unknown, the Human Genome Project estimate stands at a surprisingly low number, recently reduced from 35,000 to only 20,000–25,000 [20]. This number does not account for the potentially large number of partners that can connect to a hub protein. At the same time, to be functional and to avoid side-effects, interactions should be specific: one well documented way to obtain multiple specific interactions is through protein–protein interaction domains, such as SH2, SH3, F-box, BRCT and PDZ, to name only a few. By linking these to different chains, the effective number of partners can increase dramatically. The Pawson laboratory website (<http://pawsonlab.mshri.on.ca/>) provides a useful compilation of such domains, the parent proteins and their partners. Although the fraction of interactions occurring through such domains out of the total number of hub interactions is currently unclear, new experimental advances could assist in such endeavors [21]. Further, even if there is some over-estimation of hub interactions due to large-scale experimental procedures such as tandem affinity purification (TAP) experiments, the problem remains.

Can a single structure accommodate tens or hundreds of specific interactions?

How can we explain specific binding to multiple partners that lack interaction domains? If protein structures are in their thermodynamically most stable states and an amino acid sequence uniquely specifies a structure, how can a single structure bind to tens, or hundreds of partners? The energy landscape theory tells us that the structural ensemble will shift upon changes in the environment such as other binding events, partner concentration, pH, ionic strength and temperature [22,23]. Nonetheless, such changes are not capable of rationalizing tens or hundreds of proteins binding to the same protein, as in the case of the hubs. It can be argued that many hubs are, at least partially, disordered and as such can assume a broad range of different conformations, each of which can bind to a different partner [24,25]; yet, how can so many specific structures form from a single structure? Even if the structures, or the binding sites, are extremely flexible and can accommodate a large number of partners, this number is limited (Figure 2). To date, the observed differences in the conformational ensemble of the free state and between the free and the bound (complexed) states for a given receptor are mostly small [26,27]; as such, they do not provide an explanation for this conformational enigma. The size of hub proteins cannot resolve this problem: although hubs can be extremely large (e.g. containing hundreds or thousands of amino acids) and can have surface areas of thousands of Å² units, it is not physically feasible for a single structure to contain so many distinct binding sites [28,29]. Even if partners share binding sites [30], as currently believed to be the case, regardless of the time dimensionality of the binding events (Box 1), the question remains: how can so many (somewhat) different proteins bind at the

same site? Protein binding sites, even if promiscuous, must be able to discriminate between targets.

Considering ‘gene products’ rather than ‘proteins’ in protein–protein interaction networks

Here, we suggest that the problem does not actually exist. A single structure cannot bind hundreds of different proteins, even if it is extremely flexible or disordered. The problem lies in the confusion between ‘proteins’ and ‘gene products’ (Box 2). This proposition might appear to be semantic; but it is not: a gene can have many gene products, each with a specific amino acid sequence. Hence, in interaction networks, a node is not a single protein, but rather a collection of proteins; and an edge is an interaction involving one of these products. The interactions drawn in a map are between gene products each with a distinct structure. Such a scheme suggests that a single gene or a set of related (duplicated) genes in effect can have multiple distinct sequences, structures and interactions.

This point is important to clarify. Much effort has been invested in understanding what differentiates high-connectivity hub proteins from low-connectivity non-hub proteins [17,31]. The protein–protein interaction network of *S. cerevisiae* has shown that many of the hub interactions can be explained only partly by binding to similar proteins (or domains). Hubs are frequently multi-domain, enriched in repeats [28,29,31], and at least partly disordered [25]. Yet, to date, analyses have not revealed a property clearly differentiating hubs from non-hubs and, in particular, explaining how such hundreds of interactions can be accommodated by a single protein.

The solution would be simple if there were as many conformations as the number of partners. In such a scenario, the number of gene products would be equal to the number of interactions. The sequences of the gene products would vary to some extent from each other, with each of these sequences specifying a distinct structure (Figure 3). The differences among the conformations might not be large; however, they would be sufficient to lead to different preferred partners.

Cellular mechanisms that can lead to distinct gene products

Several well established cellular mechanisms can lead to distinct gene products; among these, alternative splicing probably leads to the largest number. In splicing, exon–intron boundaries are defined by the interactions between the spliceosome and the precursor mRNA sequence signals [32]. A large number of protein factors and small nuclear RNAs (snRNAs) assemble on the RNA in a stepwise process, forming the spliceosome. The spliceosome’s catalytic components are activated via structural rearrangement leading to intron removal. Each pre-mRNA and splice signal is recognized by specific protein factors. Alternative splicing signals lead to alternative splicing patterns. Although the prevalence of alternative splicing across organisms and its relationship to organism complexity is unclear, it is a frequent event: in the well known example of *Drosophila melanogaster Dscam*, over 38,016 isoforms are generated for the neural wiring and immune defence [33,34] (Figure 4). Although this could be an extreme case, for multi-domain proteins the number of generated

gene products can indeed be large: in the human genome, there are ~9 exons per gene on average [35]. In a simple case, in which a given exon can be included or excluded, 512 (2^9) possibilities exist. From the protein structural standpoint, this number does not account for the fact that only certain exons can be deleted: those on the protein surfaces (i.e. not in the protein cores) and those constituting independent domains. By contrast, during alternative splicing, frequently only a small number of nucleotides are spliced, often those relating to disordered loops [36] on the protein surface. In a dataset analysis of 46 differentially-spliced human genes, 80% of 75 alternatively spliced fragments were associated with fully (57%) or partially (24%) disordered regions [36]. More than half of the alternatively spliced regions contained partner interaction sites [37]. In addition, 67% of the alternatively spliced isoforms change the protein structure core significantly, likely leading to large conformation change [37]. Hence, alternative splicing expands the functional proteome and it accounts for multiple gene products in protein–protein interaction networks. Currently, a single node represents all isoforms; each with a distinct conformation and with preferred partners.

An additional diversity-generating mechanism is gene duplication and turnover of transcription start sites (TSS), which take place immediately after duplication [38]. This can occur by genes duplicated via retroposition with consequent loss of regulatory sequences and accrual of additional flanking sequences; by unequal crossing-over, which might not include the entire coding sequence, and by alternative promoter usage. Gene turnover of TSS is a frequent event: 55% of duplicate gene pairs do not retain the same coding sequences. Alternative translation initiation (ATI) can generate diversity. Here, a single mRNA leads to multiple gene products; shorter isoform products initiate from a downstream in-frame start codon, lacking the N-terminal fragment of the full-length version. ATI occurs through ribosome skipping of the first translation initiation site when it is a non-optimal Kozak sequence and moving down to the next AUG codon [39]. This mechanism was recently well documented in the K^+ channels [40,41]. Translational pausing can lead to altered protein products with similar, yet distinct, structures and functions. Ribosomal pausing events can allow a slower folding region upstream to fold to its most stable state before its association with a downstream protein segment [42]. This mechanism has been well documented in synonymous single-nucleotide polymorphisms in *MDR1* (multidrug resistance 1; also termed *ABCB1*). This process involves frequent-to-rare codon substitutions, leading to identical protein sequences [42,43]. Finally, post-translational modifications, such as different patterns of phosphorylation and glycosylation can allosterically alter the preferred conformational states by shifting the energy landscape [44,45] (Box 2); alternatively, phosphorylation can enhance/block binding sites, leading to different binding patterns [44].

Variations in multiple gene products is closely related to cancer

Variations in gene products and post-translational modifications are closely related to cancer (e.g. in the p53 family (Box 2)). Another similar example relates to the role played by the p38 MAPK pathway in oncogene-induced senescence, a tumor suppressive defense mechanism triggered upon activation of certain oncogenes in normal cells. Recent data show that the contributions of four p38 isoforms, each encoded by different genes, to senescence induction differ: p38a and p38g, but not p38 β , have an essential role in oncogenic *ras*-induced senescence [46]. Together, these provide examples of ‘proteins’ versus ‘gene

products' in the protein–protein interaction network: in each case, it is not a single structure that binds to multiple partners; rather, it is multiple gene products represented by a single node. A third example relates to BRCA1 (http://www.genenames.org/data/hgnc_data.php?hgnc_id=1100 breast cancer 1, early onset), a protein that serves many functions, including response to DNA damage, control of cell-cycle checkpoints and transcription regulation [47]. *BRCA1* and *BRCA2* intron variants located near or within consensus splicing sites appear to impact splicing fidelity [48]. For *BRCA1*, alternative splicing sites occur mostly at disordered regions [36], which could translate into larger overall changes in the protein–protein interactions it regulates. Many BRCA1 isomers are found in tumor tissues [49].

Concluding remarks: a network node can represent a large number of gene products

Based on current data, it is difficult to estimate the number of potential gene products or, consequently, the number of interactions. Nevertheless, mechanisms such as those considered above suggest that they can easily reach into the hundreds; each of these can lead to, at least, a slightly different sequence, with distinct three-dimensional structures (Figure 3). Recognizing that a network node in reality can represent a large number of products leads to greater accuracy and resolves a number of questions. Data on the fraction of the interactions mediated by protein–protein domains among the total hub interactions will be interesting, and will assist in addressing the question of how many different structures a single gene can produce. We emphasize that our discussion here excludes the immunoglobulins.

Acknowledgements

This project has been funded, in whole or in part, with federal funds from the National Cancer Institute, National Institutes of Health, under contract HHSN261200800001E. This research was supported, in part, by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

References

1. Isalan M et al. (2008) Evolvability and hierarchy in rewired bacterial gene networks. *Nature* 452, 840–845 [PubMed: 18421347]
2. Daily MD and Gray JJ (2009) Allosteric communication occurs via networks of tertiary and quaternary motions in proteins. *PLoS. Comput. Biol* 5, e1000293 [PubMed: 19229311]
3. Ghosh A and Vishveshwara S (2008) Variations in clique and community patterns in protein structures during allosteric communication: investigation of dynamically equilibrated structures of methionyl tRNA synthetase complexes. *Biochemistry* 47, 11398–11407 [PubMed: 18842003]
4. Csermely P (2008) Creative elements: network-based predictions of active centres in proteins and cellular and social networks. *Trends Biochem. Sci* 33, 569–576 [PubMed: 18945619]
5. Robinson CV et al. (2007) The molecular sociology of the cell. *Nature* 450, 973–982 [PubMed: 18075576]
6. Hopkins AL (2008) Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol* 4, 682–690 [PubMed: 18936753]
7. Keskin O et al. (2007) Towards drugs targeting multiple proteins in a systems biology approach. *Curr. Top. Med. Chem* 7, 943–951 [PubMed: 17508925]

8. Spiro Z et al. (2008) Drug-therapy networks and the prediction of novel drug targets. *J. Biol* 7, 20 [PubMed: 18710588]
9. Yosef N et al. (2009) Toward accurate reconstruction of functional protein networks. *Mol. Syst. Biol* 5, 248 [PubMed: 19293828]
10. Levy ED et al. (2009) How perfect can protein interactomes be? *Sci. Signal* 2, pe11 [PubMed: 19261595]
11. Carbonell P et al. (2009) Energetic determinants of protein binding specificity: Insights into protein interaction networks. *Proteomics* 9, 1744–1753 [PubMed: 19253304]
12. Uetz P et al. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623–627 [PubMed: 10688190]
13. Gavin A-C et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141–147 [PubMed: 11805826]
14. Krogan NJ et al. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 440, 637–643 [PubMed: 16554755]
15. Barabasi AL and Albert R (1999) Emergence of scaling in random networks. *Science* 286, 509–512 [PubMed: 10521342]
16. Aragues R et al. (2007) Characterization of protein hubs by inferring interacting motifs from protein interactions. *PLoS Comput Biol* 3, e178
17. Kim PM et al. (2006) Relating three-dimensional structures to protein networks provides evolutionary insights. *Science* 314, 1938–1941 [PubMed: 17185604]
18. He X and Zhang J (2006) Why do hubs tend to be essential in protein networks? *PLoS. Genet* 2, e88 [PubMed: 16751849]
19. Stein LD (2004) Human genome – end of the beginning. *Nature* 431, 915–916 [PubMed: 15496902]
20. Collins FS et al. (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431, 931–945 [PubMed: 15496913]
21. Boxem M et al. (2008) A protein domain-based interactome network for *C. elegans* early embryogenesis. *Cell* 134, 534–545 [PubMed: 18692475]
22. Tsai CJ et al. (1999) Folding funnels, binding funnels, and protein function. *Protein Sci.* 8, 1181–1190 [PubMed: 10386868]
23. Kumar S et al. (2000) Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci.* 9, 10–19 [PubMed: 10739242]
24. Dunker AK et al. (2005) Flexible nets - the roles of intrinsic disorder in protein interaction networks. *FEBS J.* 272, 5129–5148 [PubMed: 16218947]
25. Haynes C et al. (2006) Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS. Comput. Biol* 2, e100 [PubMed: 16884331]
26. Boehr DD and Wright PE (2008) How do proteins interact? *Science* 320, 1429–1430 [PubMed: 18556537]
27. Lange OF et al. (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* 320, 1471–1475 [PubMed: 18556554]
28. Keskin O et al. (2005) Protein-protein interactions: organization, cooperativity and mapping is a bottom-up systems biology approach. *Phys. Biol* 2, S24–S35 [PubMed: 16204846]
29. Keskin Z et al. (2008) Principles of protein-protein interactions: what are the preferred ways for proteins to interact? *Chem. Rev* 108, 1225–1244 [PubMed: 18355092]
30. Keskin O and Nussinov R (2007) Similar binding sites and different partners: Implications to shared proteins in cellular pathways. *Structure* 15, 341–354 [PubMed: 17355869]
31. Ekman D et al. (2006) What properties characterize the hub proteins of the protein-protein interaction network of *Saccharomyces cerevisiae*? *Genome. Biol* 7, R45 [PubMed: 16780599]
32. Hertel KJ (2008) Combinatorial control of exon recognition. *J. Biol. Chem* 283, 1211–1215 [PubMed: 18024426]
33. Meijers R et al. (2007) Structural basis of Dscam isoform specificity. *Nature* 449, 487–491 [PubMed: 17721508]

34. Schmucker D (2007) Molecular diversity of Dscam: recognition of molecular identity in neuronal wiring. *Nat. Rev. Neurosci* 8, 915–920 [PubMed: 18026165]
35. Sakharkar MK et al. (2004) Distributions of exons and introns in the human genome. *In Silico. Biol* 4, 387–393 [PubMed: 15217358]
36. Romero PR et al. (2006) Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc. Natl Acad. Sci. USA* 103, 8390–8395 [PubMed: 16717195]
37. Yura K et al. (2006) Alternative splicing in human transcriptome: functional and structural influence on proteins. *Gene* 380, 63–71 [PubMed: 16872759]
38. Park C and Makova K (2009) Coding region structural heterogeneity and turnover of transcription start sites contribute to divergence in expression between duplicate genes. *Genome. Biol* 10, R10 [PubMed: 19175934]
39. Honore E (2008) Alternative translation initiation further increases the molecular and functional diversity of ion channels. *J. Physiol. (London)* 586, 5605–5606 [PubMed: 19043120]
40. Simkin D et al. (2008) Control of the single channel conductance of K(2P)10.1 (TREK-2) by the amino-terminus: role of alternative translation initiation. *J. Physiol. (London)* 586, 5651–5663 [PubMed: 18845607]
41. Thomas D et al. (2008) Alternative translation initiation in rat brain yields K(2P)2.1 potassium channels permeable to sodium. *Neuron* 58, 859–870 [PubMed: 18579077]
42. Tsai CJ et al. (2008) Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *J. Mol. Biol* 383, 281–291 [PubMed: 18722384]
43. Kimchi-Sarfaty C et al. (2007) A “silent” polymorphism in the MDR1 gene changes substrate specificity. *Science* 315, 525–528 [PubMed: 17185560]
44. Latzer J et al. (2008) Conformational switching upon phosphorylation: a predictive framework based on energy landscape principles. *Biochemistry* 47, 2110–2122 [PubMed: 18198897]
45. Tsai C-J et al. (2009) Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. *Mol. Biosyst* 5, 207–216 [PubMed: 19225609]
46. Kwong J et al. (2009) p38alpha and p38gamma mediate oncogenic ras-induced senescence through differential mechanisms. *J. Biol. Chem* 284, 11237–11246 [PubMed: 19251701]
47. Ashworth A (2004) Refocusing on BRCA1. *Nat. Cell Biol* 6, 916–917 [PubMed: 15459717]
48. Tommasi S et al. (2008) Molecular and in silico analysis of BRCA1 and BRCA2 variants. *Mutat. Res* 644, 64–70 [PubMed: 18694767]
49. Orban TI and Olah E (2003) Emerging roles of BRCA1 alternative splicing. *Mol. Pathol* 56, 191–197 [PubMed: 12890739]
50. Tuncbag NK et al. (2009). Inferring time dimensionality in the protein-protein interaction networks by integrating structures: the p53 example. *Mol. Biosyst* (in press) doi:10.1039/B905661K
51. Han JD et al. (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* 430, 88–93 [PubMed: 15190252]
52. de Lichtenberg U et al. (2005) Dynamic complex formation during the yeast cell cycle. *Science* 307, 724–727 [PubMed: 15692050]
53. Tozluoglu M et al. (2008) Cataloging and organizing p73 interactions in cell cycle arrest and apoptosis. *Nucleic. Acids Res* 36, 5033–5049 [PubMed: 18660513]
54. Marcel V and Hainaut P (2009) p53 isoforms - a conspiracy to kidnap p53 tumor suppressor activity? *Cell Mol. Life Sci* 66, 391–406 [PubMed: 18854945]
55. Graupner V et al. (2009) Functional characterization of p53beta and p53gamma, two isoforms of the tumor suppressor p53. *Cell Cycle* 8, 1238–1248 [PubMed: 19305147]
56. Kruse JP and Gu W (2008) Snapshot: p53 posttranslational modifications. *Cell* 133, 930–30 e1 [PubMed: 18510935]
57. Di Lello P et al. (2006) Structure of the Tfb1/p53 complex: Insights into the interaction between the p62/Tfb1 subunit of TFIIH and the activation domain of p53. *Mol. Cell* 22, 731–740 [PubMed: 16793543]

Box 1.**Time dimensionality in protein–protein interaction networks**

Recently, Tuncbag *et al.* [50] proposed a new concept toward inferring time dimensionality in protein–protein interaction networks. Including the notion of time within a protein–protein structural network [50] is an immense asset. Incorporating time dimensionality would transform static network node-and-edge maps into dynamic processes and in this way help in the understanding of the regulation of cellular pathways. However, incorporation of time into the networks is difficult. Han *et al.* [51] divided hub proteins into ‘date’ and ‘party’ hubs on the basis of mRNA expression profiles. If the expression profiles of the hub protein and (most of) its partners are well correlated, the hub is considered a party hub; if they are not, the hub is labeled date hub. Party interactions are permanent; date interactions are transient. de Lichtenberg *et al.* [52] mapped the expression data onto the interaction network to obtain the dynamics of complex formation during the yeast cell cycle. Kim *et al.* [17] added a structural perspective by subdividing the hubs according to the number of interfaces that they have to ‘single-interface’ and ‘multi-interface’ hubs. In principle, if robust, time series of mRNA expression array data could be a powerful way of obtaining time dimensionality of protein–protein interactions in an immensely fluctuating cellular environment; however, currently there are no such available large-scale data. Tuncbag *et al.* [50] showed that the use of state of the art structural comparison algorithms, protein–protein interface datasets and appropriate filters, assists in the prediction of which interactions can occur simultaneously and which are mutually excluded. For hub proteins, which can bind hundreds of partners, information relating to which proteins can be bound at the same time and which cannot, should be extremely useful in the understanding of cellular regulation at the molecular level. Predictions are compared with experimental protein interaction data. We expect that an elucidation of the relative timing, order of events, and time-scales of the interactions can be obtained through the combined use of process diagrams or molecular interaction maps [53] that include contingencies representing effects such as inhibition and stimulation, larger scale multi-molecular complexes, and expression array time. From the perspective presented here, a party hub is expected to interact with fewer partners than a date hub.

Box 2.**p53, p63 and p73 isoforms**

p53, p63 and p73 isoforms are examples of multiple gene products in protein–protein interaction networks which display post-translation modification, alternative splicing and transcription initiation. The tumor suppressor p53 has multiple, important cellular functions such as those related to apoptosis, downstream cell cycle regulation via p21, and control of tumor angiogenesis by cyclin-dependent kinases. Somatic mutations in *TP53* are the most common genetic mutation in human cancer. p53 was long considered to be the only protein expressed by *TP53* [54]; recently, however, alternative transcripts have been observed, in both mouse and human, with synergistic or antagonistic roles in the modulation of p53 suppressive activity. Recent studies [55] indicated that, despite the efficient nuclear expression of all isoforms, only wild type p53 (p53 wt) modulates the induction of apoptosis; only p53 wt assembles into a transcription-competent oligomeric complex; and only p53 wt translocates to mitochondria upon stress induction. By contrast, C-terminally truncated isoforms, such as p53 β and p53 γ , do not modulate the apoptotic function of p53 wt; they are unable to associate with p53 wt and cannot bind to p53 DNA recognition sequences. Similarly, the dominant-negative function of transactivation-deficient deletion mutant 133p53 is lost in the 133p53 β variant. Yet, the alternatively spliced C-terminus protects p53 β and p53 γ from MDM2-mediated proteasomal degradation and strongly impairs their binding to this negative regulator. These data suggest that the C-terminal tail is essential for multiple layers of p53 regulation and function, implying that alternatively spliced C-terminal mutants correspond to multiple gene products; not simply to a single protein with a single structure. p53 isoforms undergo extensive post-translational modifications [56], turning on/off specific interactions based on the phosphorylation states (Figure I). For example, phosphorylation of S46 and T55 leads to marked enhancement in p53 binding to p62 and TFB1[57]. Taken together, the pyramidal expansion of p53 products from transcriptional variation, post-translational modification and cofactor binding permit p53 to regulate hundreds of genes and proteins.

TP53 belongs to a family containing two other members, *TP63* and *TP73*, which are expressed as multiple isoforms and whose protein products have roles in development and differentiation. The expression of different isoforms results from both multiple promoter usage and alternative splicing. The different pattern of binding is well illustrated in the p73 interaction map [53]. Currently, seven C-terminally alternatively spliced truncation types have been identified for p73. N-terminally truncated isoforms are formed by an alternative promoter in intron 3; they can have C-terminal truncations. Notably, N-terminal deletion mutants display deficient transactivation activity.

Figure 1.

A simple diagram illustrating a protein–protein interaction network. (a) Each node represents a protein and each linkage between nodes represents a direct specific protein–protein interaction (indirect interactions via mediating proteins are excluded). In this simplified network, proteins are illustrated as having various numbers of interacting partners. A node with more than ten interactions (in reality these can number in the hundreds) with other nodes is considered a hub protein. Linkages of a particular node can be mutually exclusive (i.e. cannot exist at the same time) if their partners use the same interaction interface in that node protein. (b) A hub protein with three distinctive regions (in reality there could be more than eight) recognizing its partners via a specific binding motif. (c) At the same time, at each binding region, a hub protein can have several distinct conformations, each dedicated to a specific protein–protein interaction with its partners. (d) The conformation of a specific protein–protein interaction should include all minor conformational changes due to ‘fine tuning’ via thermal vibration.

Figure 2.

How can a hub protein with a single structure interact with many different partners? **(a)** A single conformation at a binding region can interact with many different partner proteins through a combinatorial domain linkage to a common binding motif. Here, we show only three out of a total of 15 possible different proteins. In each protein, two different domains were added to a common binding motif from a pool of six distinct domains. In principle, the number of partners can be multiplied by the number of newly populated conformations at the same binding region. **(b)** An allosteric binding effect. The conformation at the binding region can change to accommodate a new specific interaction with another binding motif. **(c)** Even without the help of allostery, a disordered protein already has a pool of equally populated conformations. This is due to the inherent lower barriers encountered when a disordered protein switches from one specific conformation to another as compared with a folded protein.

Figure I.

p53 is one of the most connected nodes in either the protein–protein interaction network or the gene regulation network. The central circle represents both wild type p53 and isoforms, the light brown circles are phosphorylation events required for turning on/off specific p53–protein interactions. The grey circles represent proteins interacting with p53. Note that the phosphorylation of Ser15 and Ser20 blocks the p53–MDM2 interaction.

Figure 3.

A simple diagram illustrating a hub protein (represented by one gene) with many distinct specific conformations that could originate from different structures. **(a)** A pre-mRNA with four exons (coloured bars) and three introns (black bars) can produce three different mRNAs via an exon skipping mechanism (one of the major forms of alternative splicing), which in turn correspond to three different protein structures. **(b)** A gene whose mRNA undergoes significant translational pausing due to multiple silent mutations can generate a protein structure different from its 'normal' structure. **(c)** A post translational modification (e.g. phosphorylation) can allosterically or directly alter the protein structure to create a new conformation.

Figure 4.

Alternative splicing generates large number of isomers with different protein–protein interaction possibilities. (i) *Dscam* has 24 exons; exon 4 has 12 variants, exon 6 has 48 variants, exon 9 has 33 variants, and exon 17 has two variants. The combination of exons 4, 6, and 9 leads to 19,008 possible isoforms with different extracellular domains (due to differences in Ig2, Ig3 and Ig4). With two different transmembrane domains from exon 17, the total possible protein products could reach 38,016 isoforms. (ii) These isoforms have different specific interactions due to structural variability. (iii) For example, the crystal structures of two isoforms (first four domains only) revealed large variation in their binding epitopes, resulting in totally different dimer forms. Red ribbons represent the structure of isoform D9.9 which is the product of variant 9 of exon 4 and variant 9 of exon 6 (PDB code 2V5R), the green ribbons represents isoform D1.34, which is the product of variant 1 of exon 4 and variant 34 of exon 6 (PDB code 2V5S). The figure in the lower left corner is the superimposition of the two monomers of D9.9 and D1.34 with the differences in the large Ig3 domain loop highlighted. The figure in the lower right corner is the superimposition of two dimers of D9.9 and D1.34, with a completely different interface and dimer conformations.