Robustness of the p53 network and biological hackers

Lewis Dartnell[a,1], Evangelos Simeonidis[b,1], Michael Hubank[c], Sophia Tsoka[d], I. David L. Bogle[a,b], Lazaros G. Papageorgiou[b,*]

[a] *Centre for Mathematics and Physics in the Life Sciences and Experimental Biology (CoMPLEX), University College London (UCL), London NW1 2HE, UK*
[b] *Centre for Process Systems Engineering, Department of Chemical Engineering, University College London (UCL), London WC1E 7JE, UK*
[c] *Molecular Haematology and Cancer Biology Unit, Institute for Child Health (ICH), University College London (UCL), London WC1N 1EH, UK*
[d] *Computational Genomics Group, Research Program, The European Bioinformatics Institute, EMBL Cambridge Outstation, Cambridge CB10 1SD, UK*

Received 3 December 2004; revised 16 March 2005; accepted 23 March 2005

Available online 5 May 2005

Edited by Robert B. Russell

Abstract The p53 protein interaction network is crucial in regulating the metazoan cell cycle and apoptosis. Here, the robustness of the p53 network is studied by analyzing its degeneration under two modes of attack. Linear Programming is used to calculate average path lengths among proteins and the network diameter as measures of functionality. The p53 network is found to be robust to random loss of nodes, but vulnerable to a targeted attack against its hubs, as a result of its architecture. The significance of the results is considered with respect to mutational knockouts of proteins and the directed attacks mounted by tumour inducing viruses.

Keywords: p53 cell cycle and apoptosis control network; Network robustness; Protein knockout; Linear programming; Tumour inducing virus

1. Introduction

1.1. The p53 network

In multicellular organisms (metazoans), cellular proliferation is tightly regulated. Cells accumulate in a co-ordinated way during growth or repair, and undergo programmed cell death (apoptosis) when genetically damaged, virally infected or the developmental program requires it. Proliferation is regulated via cyclical activation of different cyclin-dependent kinases (CDKs), which mediate the temporal activation of cell growth, DNA synthesis and cell division. Apoptosis is triggered in response to specific signals, and the cell is destroyed when a cascade of proteases (the caspases) are activated. Failure of this control system, leading to either unregulated proliferation or unnecessary apoptosis, is causative of both tumourigenesis and developmental diseases.

All organisms control progression through the cell cycle, and can respond to cellular stress by activating cell cycle checkpoints and repairing damaged components if necessary. In addition, multicellular organisms have evolved the ability to trigger apoptosis in cases where risk to the organism is unacceptably high. In response to stress, a metazoan cell must decide between continued (or resumed) progression through the cell cycle or initiation of apoptosis. This decision is mediated by a protein-interaction network, at the centre of which lies the p53 protein. p53 is found only in metazoan cells, and combines protein interaction domains, regulatory domains and a sequence specific DNA recognition domain that allow the integration of intra- and intercellular signals with gene transcription [1]. Under normal conditions, p53 is turned over rapidly by proteolysis and is inactive. Cellular stress signals result in the stabilization of p53 so that it rises in concentration to a level where it can activate transcription of its target genes. Depending on the circumstances (for example cell type and nature and strength of the stress signal), gene transcription resulting from elevated p53 levels produces responses including pausing of cell cycle, DNA repair, permanent arrest of replication, or apoptosis. p53 is activated by both intrinsic and extrinsic stress signals, including DNA damage (e.g. from ionizing radiation), mitotic spindle damage, aberrant growth signals, hypoxia, ribonucleotide depletion, and loss of cell adhesion [2], many of which are indicators of tumourigenesis. The importance of the p53 response network in the prevention of cancer is striking, and mutations reducing p53 activity are present in over 50% of human tumours [3].

1.2. Network architectures

Graph theory is a branch of mathematics used to analyse complex networks of nodes and connections. Initial work focused on two main types of graphs; regular and random. The connections in a regular graph are very strictly ordered with all nodes having the same degree, k (number of connections to other nodes), much like the chemical bonds in a crystal lattice. In a random network connections are placed between any two nodes with a given probability.

The structure of a network with N nodes is often summarised by plotting k against the probability distribution

*Corresponding author. Fax: +44 20 7383 2348.
E-mail address: l.papageorgiou@ucl.ac.uk (L.G. Papageorgiou).

[1] These authors equally contributed to this work.

0014-5793/$30.00
doi:10.1016/j.febslet.2005.03.101

function, $P(k)$ (the number of nodes with k connections, divided by the total number of nodes in the network). The plot of $P(k)$ for a random graph follows a Poisson distribution peaking at the average value of k. For a regular network the plot is a spike as all nodes have the same connectivity. Counting the minimum number of connections that must be followed to traverse from one node, i, to another, j, yields the path length for that pair, l_{ij}. Within a random graph, connections can 'short-cut' across the entire network, and so path lengths are typically much shorter than in a regular graph [4]. One global metric of the structure of a network is its diameter, D; the average path length among all nodes. It is defined as:

$$D = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j \geqslant i+1}^{N} l_{ij} \quad (1)$$

Watts and Strogatz [5] described the properties of a third class of graphs. Their 'small-world' networks combine the small diameter of random graphs with the high local connectivity of regular graphs. In addition to this small-world property, many networks with no pre-designed architecture that grow and evolve over time have a characteristic pattern of connectivity. $P(k)$ decays as a power law – the vast majority of nodes have only a few connections, but there are several hubs that are very highly connected. Unlike regular or random graphs there is no characteristic degree of connectivity, and so such networks are termed 'scale-free'. In recent years, a great number of networks have been shown to be scale-free, including the Internet [6], social interactions [7], neural networks [8], ecological food webs [8], metabolism [9], protein–protein interactions [10], and gene transcription regulation networks [4]. One explanation for the occurrence of this structure is that of network growth through preferential attachment of additional nodes. New nodes are added and connected to existing nodes with a probability proportional to their current connectivity, with hubs being created by a positive-feedback 'rich getting richer' process. Barabási and Oltvai [4] describe how this process might operate in protein–protein interaction networks through gene duplication.

1.3. Network robustness

Much research has been conducted into the robustness of networks, that is, their ability to remain relatively undisrupted in the face of perturbation. Robustness can be defined, in topological terms, as the remaining communication ability within a network as nodes or connections are removed. Real networks also perform a function, be it electricity distribution or genetic regulation, but modelling a complex network's performance is often prohibitively difficult. Network navigability is a necessary (although not sufficient) prerequisite for adequate function, and so the diameter of the network is taken as an acceptable proxy [7].

Either individual connections, or entire nodes can be removed from a network, with the latter having a greater impact. There are two main modes of attack upon the nodes of a network – either removed at random, or the preferential targeting of the hubs. A random graph responds identically to both random and directed attacks as its connectivity is homogenous – the majority of its nodes have roughly the same number of links, approximately equal to the network's average degree. Upon successive deletion of nodes the diameter increases monotonically until a critical threshold fraction has been exceeded, and the network undergoes a phase transition as it disintegrates into isolated fragments. In contrast, a scale-free network is relatively immune to random node failure, but vulnerable to a targeted onslaught [4].

2. Methods

A model of the p53 network, a large module of the entire metazoan protein interaction network, was constructed. The model was then subjected to both random and directed modes of attack, and its changing diameter studied. The objective was to analyze how the network behaves in response to the stochastic protein knockouts from mutation during tumourigenesis, and also in response to a targeted attack.

2.1. Raw data

There are over 35 000 published articles relating to p53, its interactions, its functions, and the consequences of its inactivation. These articles describe p53 function in multiple organisms, multiple cell types within those organisms, and under a wide range of different situations within those cells. The result is a bewildering volume of information relating to p53 interactions, the relative merits of which can be extremely difficult to judge. As a result, very few studies have attempted to fully connect the network in any meaningful way. Kohn [11] performed an extensive literature review and presented an annotated molecular interaction map of proteins involved in mammalian cell cycle progression and checkpoints, DNA repair, and apoptosis. The molecular maps of these processes all feature p53 prominently, and although the connections are incomplete and inevitably contain inaccuracies, the majority of the described interactions are experimentally validated and well understood. It is unlikely that a few false-positives or negatives would drastically alter the architecture of the network or the calculations of its diameter.

Although there is a high degree of modularity within interaction networks, the dissociation of the ''p53 network'' will follow largely arbitrary borders. Our study therefore followed the same boundaries selected by Kohn, yielding a network containing 104 nodes and 226 unique connections. A representation of this was constructed using the Pajek program for large network analysis (http://vlado.fmf.uni-lj.si/pub/networks/pajek/), as shown in Fig. 1. The vast majority of the nodes can be seen to be poorly connected within the network, whereas very few of the nodes are hubs with a high centrality.

All interactions were assumed to be mutual. This is a valid assumption for proteins reciprocally binding in a complex, but transcription regulation events are directional. Such interactions account for only around 5% of the total described by Kohn. Although the linear programming algorithm used [12] is capable of dealing with directionality, the entire map was taken to be undirected to simplify the analysis. Several nodes such as ssDNA (single stranded DNA), are not proteins, but are nonetheless crucial objects that interact within the p53 control network and so are included in this study.

2.2. Minimum path lengths

Linear programming (LP) is an extensively used technique for the solution of problems involving the optimum allocation of limited resources to competing demands. For this study, an LP algorithm is used to calculate the shortest paths in the p53 network. This technique was previously proposed to study minimum pathway distances in *E. coli* small molecule metabolism [12]. The algorithm is capable of finding in a single pass the shortest path lengths from a source protein, i^*, to all other proteins in the network. The notation used in the mathematical formulation is as follows:

Indices

i,j: proteins

Parameters

C_{ij}: 1 if there is a connection (interaction) between i and j; 0 otherwise

T: large number

Continuous variables

l_i: path length from i^* to i

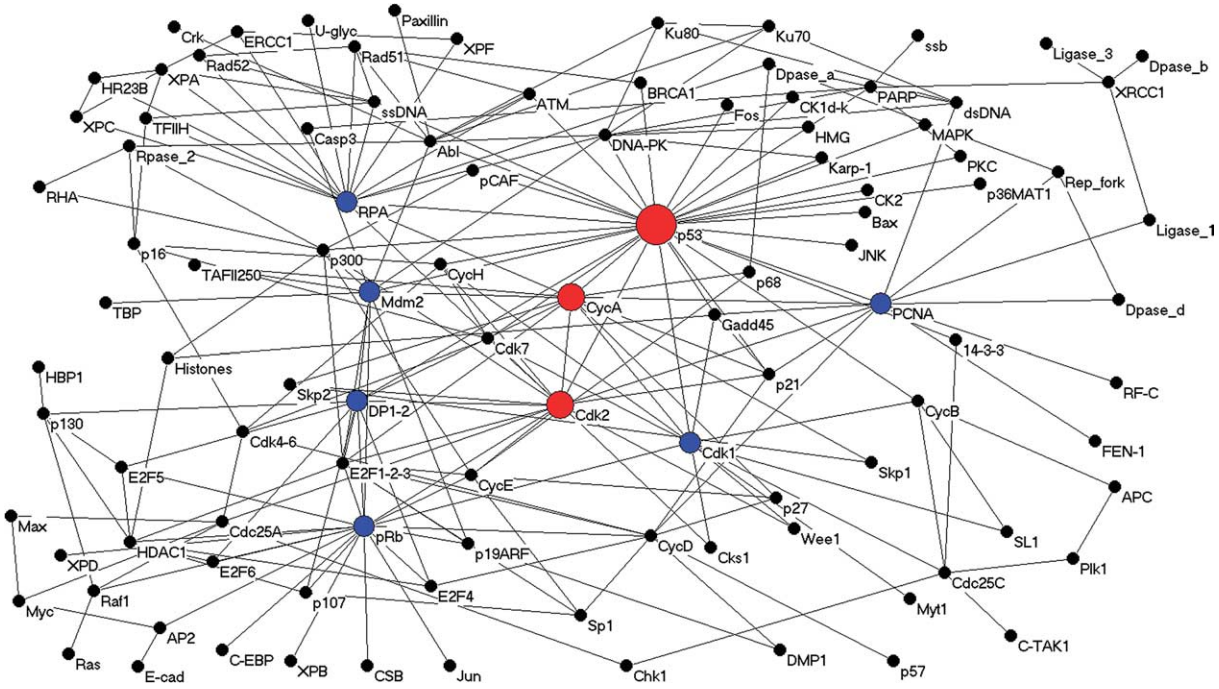

Fig. 1. Visualization of the p53 network (based on data from Kohn [11]).

Each protein is systematically set as the source, i^*, and the algorithm finds the shortest path to all other proteins by solving the following LP optimisation model:

$$\text{maximise} \quad \sum_i l_i \tag{2}$$

subject to:

$$l_j \leqslant l_i + 1 \quad \forall(i,j)\colon C_{ij} = 1 \tag{3}$$

$$l_{i^*} = 0 \tag{4}$$

$$l_i \geqslant 0 \tag{5}$$

Constraints (3) incorporate information related to network connectivity, facilitated by the use of parameter C_{ij}. Constraint (4) assigns the initial value of zero to protein i^* to denote it as the source protein, while constraint (5) requires all l_i variables to take positive values. Finally, unbounded solutions can be avoided by adding:

$$l_i \leqslant T \quad \forall i \tag{6}$$

It should be noted that if l_i equals T at the final solution then it can be concluded that there is no path connecting the i^* source node to protein i. In this case, the diameter computation would involve infinite path lengths, and so the arbitrarily large value of T is substituted to allow calculation.

The average path length (APL) for a protein is the mean of the shortest paths between this and all other nodes in a network, summarising the propinquity of a node to every other. Fell and Wagner [13] calculated this metric for metabolites in *E. coli* core metabolism. Similarly, this study uses the average path length for proteins in the p53 network, as a global measure of a node's centrality.

2.3. Network attacks

The survivability of the p53 network in the face of both a directed or random attack against its nodes was examined. At each stage of the attack an additional protein was knocked-out and the diameter recalculated until all 104 had been removed, the sequence specified either by a random permutation or in the case of the directed attack by rank order of centrality. The random attack was repeated 100 times, and the diameter at each step averaged across all runs. Here, the removal of a node destroyed all the connections it possessed, but an attack against only one connection at a time is also possible.

When a hub is knocked-out, nodes may become isolated from the rest of the network. This produces nodes with no navigable route to each other, in which case the path lengths are set equal to the arbitrarily large number T, given the value of 100 in this study. Tests were performed for other values of T as well, and it was demonstrated that the behaviour of the network is unaffected by this choice.

3. Results and discussion

Fig. 2 shows the relationship between connectivity, k, and the probability distribution function, $P(k)$. The number of nodes with a given connectivity decay logarithmically as connectivity increases. This power law is a defining feature of a

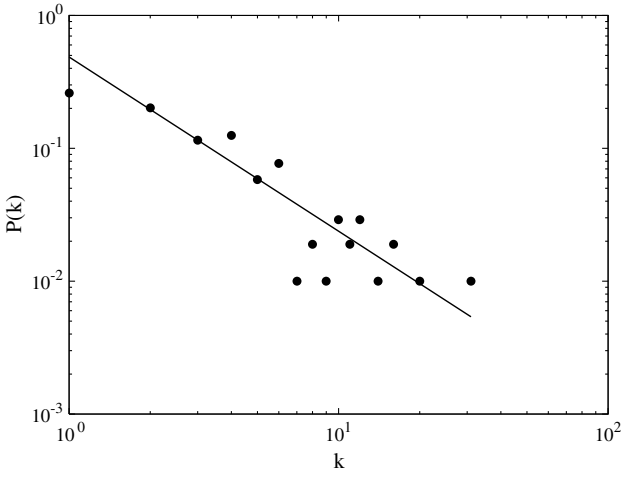

Fig. 2. The power-law relationship between k and $P(k)$.

scale-free graph, so it is possible that the p53 network also possesses such architecture.

For the calculation of the average path lengths and diameter of the network, the LP algorithm was solved with GAMS [14], using the CPLEX 6.5 solver algorithm, and run on a RS6000 workstation. Table 1 shows the nodes with lowest APLs, with hubs being defined as those proteins with the very lowest scores.

Fig. 3 shows the plot of network diameter over the first 30 knockouts with nodes removed in either a random attack, or one directed against the hubs (nodes knocked-out in rank order of APL). The diameter of the p53 network under a random attack increases very slowly. Thanks to its architecture, the majority of nodes in the network are poorly connected and therefore their removal has a very small effect on network navigability. Hub nodes are uncommon and so they are rarely hit. The p53 network is shown here to be resistant to a random pattern of attack, which equates to robustness to mutational perturbation.

Under a directed onslaught, however, network communication fails as the diameter of the network rapidly degenerates. The result of knocking out the first hub protein, p53, is an increase in network diameter of over five-fold; from 3.1 to 16.1. The loss of no other protein has such a devastating effect: removal of the next four hubs produces only a further doubling of the diameter. After the 24th knockout the diameter levels off: the directed attack has removed all of the hubs, and consequently the majority of routes between proteins. The network has shattered into isolated subclusters, and most of the path lengths between protein pairs are designated as 100, therefore further knockouts can damage the network no further.

To assess the results statistically, the standard normal deviate or Z-score is used, which measures the distance of any value from the mean of a population in standard deviation units. The results of the directed attack performed on the p53 network (see Fig. 3) are compared to the results of the same pattern of attack on 100 random networks. For each random network, 104 protein nodes were linked through random generation of 226 edges; the mean diameter ($\bar{D}$) and the standard deviation (σ) from 100 random networks, and the p53 network's diameter (D) were used for Z-score calculation. Note that diameter values were first normalised to compensate for the fact that the p53 network has a lower initial diameter from any random network, due to its scale-free nature.

$$Z = \frac{D - \bar{D}}{\sigma} \tag{7}$$

Fig. 4 presents statistical analyses using Z-scores between the p53 network and random networks. Z-scores indicate how far and in what direction each item deviates from the random mean. Z-score values greater than 3 are typically considered to be significant. As can be seen in Fig. 4, our results differ considerably from random; thus indicating that trends observed cannot be attributed to chance.

The plot of network degeneration under directed attack (Fig. 3) shows some interesting features. There are several small plateaux where diameter is temporarily stable. For example, navigability barely alters between the 2nd and 3rd knockouts (CDK2 and Cyclin A, respectively). These two proteins bind together to allow progression through a cell cycle checkpoint and activation of the DNA replication machinery. They thus bind to a large number of the same proteins, and so within this model are largely redundant – it is not until the second one is knocked out that routes between certain nodes are lost and the diameter jumps up. This behaviour under attack is an artefact of the nature of the model, however. In reality, CDK2 and

Table 1
The 30 best-connected nodes in the p53 network, in order of ascending Average Path Length (APL)

APL	Proteins
1.9	p53
2.1	Cdk2
2.2	CycA
2.3	Cdk1, Mdm2, DP1-2, pRb
2.4	PCNA, RPA
2.5	DNA-PK, p21, p300, E2F1-2-3, Cdk7, CycH
2.6	Abl, Gadd45
2.7	CycB, CycD, CycE, PARP, ATM
2.8	ssDNA, Cdc25A, 14-3-3, pCAF, PKC
2.9	HMG, Karp-1, BRCA1

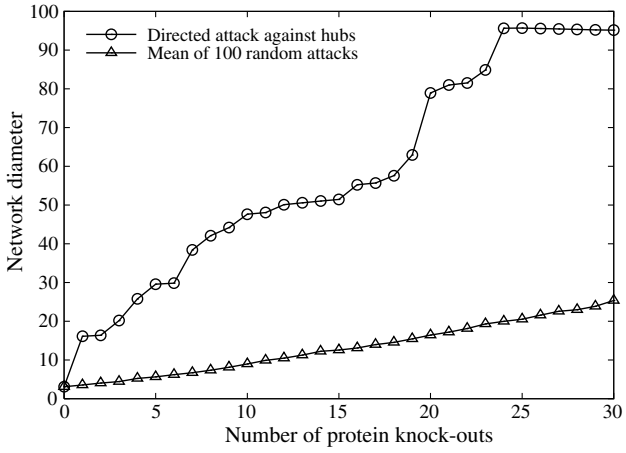

Fig. 3. Degeneration of network diameter when nodes are knocked out in either a random pattern, or in a directed attack against the hubs.

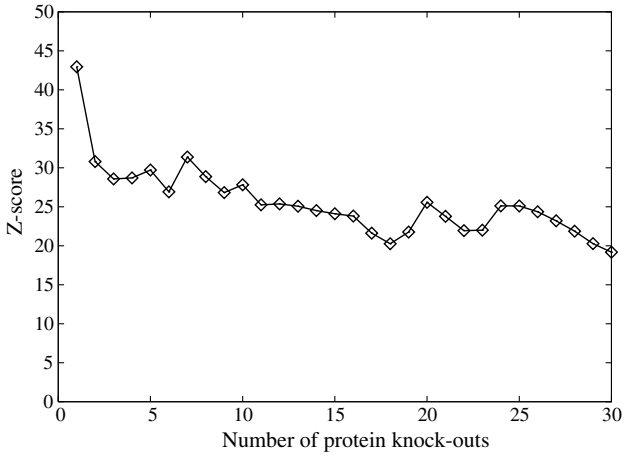

Fig. 4. At each number of knock-outs, the Z-score of the diameter of the network is plotted. The mean and standard deviation used for the estimation of the Z-score are those of random networks with 104 nodes and 226 unique connections.

Cyclin A bind together as a complex and so need each other in order to function. Knockout of either results in a loss of function of the other (and thus disappearance of its connections), in other words, such nodes in the p53 network are not strictly independent. The model also assumes bidirectionality of connections, a non-dynamic presence of proteins in the cell (i.e. the temporal component of protein concentrations is ignored), and all connections are given an equal weighting. When more complete data on protein dependencies, protein complexes, temporal fluctuations, and relative importance of interactions become available, these factors should be incorporated into models such as this one.

3.1. Biological hackers

The complete set of proteins involved within the p53 network and connections among them is still being mapped out, but the power-law relationship demonstrated in Fig. 2 suggests that it may posses a scale-free structure. The p53 network shows a similar response under attack to scale-free networks such as the Internet; it is robust against a random attack as most of the protein knockouts will have negligible impact on the global integrity of the network. This reliance on highly connected nodes, however, renders the network vulnerable to a directed attack. The most important nodes are selectively targeted and the diameter of the p53 network rapidly degenerates. A similar result was obtained on simulated attacks on the Internet, which was found to be robust to random server failures, but vulnerable to the activities of hackers deliberately targeting the hubs so as to wreak maximal havoc [7]. The question arises as to whether it is biologically possible to orchestrate a directed attack against the p53 network.

Such a threat does in fact exist in nature, operating not at a genetic level, but against the translated proteins. DNA tumour-inducing viruses (TIVs) increase their replication rate and survival with an armoury of proteins that suppress the normal apoptotic infection response, short-circuit the cell cycle into continually synthesizing viral DNA, and force the cell into a stealth mode to evade immune system surveillance. Table 2 summarises the oncoproteins produced by adenoviruses, and the effect of their inhibition of target cellular proteins (data from [15–17]).

The two most common targets, p53 and pRb, are also two of the most central proteins in the network, with APLs of 1.9 and 2.3, respectively. Knocking out multiple downstream effector proteins would have the same effect, but as this study suggests it is much more efficient to selectively remove the hubs. This is especially important for viruses, as their genome is often optimized for rapid replication and cannot afford the information cost of coding many oncoproteins. It is not advantageous for the TIVs to completely destroy the p53 network either (as then no DNA replication or cell division would occur); they need only disable regions that halt cell cycle progression or trigger apoptosis. It is conceivably for this reason that not all of the target proteins are hubs (although none have an APL greater than 3.2), but their removal disables specific functions of the p53 network. The final column in Table 2 shows the calculated diameter after the targeted proteins of the third column have been knocked out using the LP algorithm. The TIV directed strikes are effective at disrupting communication within the p53 network, but do not increase the diameter so much that the network shatters and function fails completely. TIVs thus behave like biological hackers, targeting their attack against some of the p53 network hubs and so exploiting the weakness in its architecture.

4. Conclusions

The p53 cell cycle and apoptosis control network is inherently robust to random knockouts of its proteins, which signifies resilience against mutational perturbation provided by the structure of the network itself. This robustness against mutations, however, gives the network an Achilles Heel, as the reliance on highly-connected nodes makes it vulnerable to the loss of its hubs. Evolution has produced organisms that exploit this very weakness in order to disrupt the cell cycle and apoptosis system for their own ends: tumour inducing viruses target specific proteins to disrupt the p53 network, and this study has identified these same proteins as the network hubs. Although TIVs have previously been likened to 'biological hackers', here we show why the TIV attack is so effective – TIVs target a specific vulnerability of the network that can be explained in terms of network architecture. A Z-score analysis of the results has demonstrated that our findings differ considerably from random and cannot be attributed to chance.

From the computational perspective, we display the effectiveness of the algorithm in analysing the properties of a large protein interaction signalling network. The algorithm

Table 2
Tumour inducing viruses, the nodes in the p53 network their oncoproteins target, and the extent of damage inflicted on the network by those knockouts

TIV	Viral protein	Host protein	Effect	Diameter after KO
Adenovirus	E1a	pRb	Apoptosis evasion	24.98
	E1b/55Kd	P53	Apoptosis evasion	
Coxsackie	unknown	cyclin D1	Transcription/reactivation	5.00
HCMV[a]	pp71	pRb, p107, p130	Transcription/reactivation	14.37
HPV[b] 16/18	E6	P53	Apoptosis evasion	27.07
	E7	pRb, p107, p130	Transcription/reactivation	
HSV[c]	ICP0[d]	DNA-PK	Transcription/reactivation	3.12
SV40[e]	Lg T Ag[f]	pRb, p53	Apoptosis evasion	24.98

[a]Human cytomegalovirus.
[b]Human papillomavirus.
[c]Herpes simplex virus.
[d]Infected cell protein.
[e]Simian virus 40.
[f]Large T antigen.

(previously applied to the analysis of the *E. coli* small molecule metabolism network [12]) has thus been proven to be a valuable analysis tool for complex biological networks.

Finally, the connectionist network presented here is a first-level model of the p53 cell cycle and apoptotic control network with a specific and clearly-defined function. The fact that we can represent and test the p53 network offers the future possibility to attach directions and strength values to the connections as more biological data become available, in order to make accurate predictions about the importance of individual nodes and edges. This will allow frameworks like the one presented to be used in comparative analyses of how and why the variable dynamic network components operate under different evolutionary and cell type conditions. This paper represents the first step in this exciting process.

Acknowledgments: LD is financially supported by an EPSRC/MRC Doctoral Training Centre studentship. ES was financially supported by EPSRC (Award No. 00319001) and the Centre for Process Systems Engineering. ST acknowledges funding from the UK Medical Research Council through a Special Training Fellowship in Bioinformatics.

References

[1] Vogelstein, B., Lane, D. and Levine, A.J. (2000) Surfing the p53 network. Nature 408, 307–310.
[2] Robles, A.I., Linke, S.P. and Harris, C.C. (2002) The p53 network in lung carcinogenesis. Oncogene 21, 6898–6907.
[3] Haupt, S., Berger, M., Goldberg, Z. and Haupt, Y. (2003) Apoptosis – The p53 network. J. Cell Sci. 116, 4077–4085.
[4] Barabási, A.L. and Oltvai, Z.N. (2004) Network biology: Understanding the cell's functional organization. Nat. Rev. Genet. 5, 101–113.
[5] Watts, D.J. and Strogatz, S.H. (1998) Collective dynamics of 'small-world' networks. Nature 393, 440–442.
[6] Albert, R., Jeong, H. and Barabási, A.L. (1999) Internet – Diameter of the World-Wide Web. Nature 401, 130–131.
[7] Albert, R., Jeong, H. and Barabási, A.L. (2000) Error and attack tolerance of complex networks. Nature 406, 378–382.
[8] Strogatz, S.H. (2001) Exploring complex networks. Nature 410, 268–276.
[9] Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N. and Barabási, A.L. (2000) The large-scale organization of metabolic networks. Nature 407, 651–654.
[10] Jeong, H., Mason, S.P., Barabási, A.L. and Oltvai, Z.N. (2001) Lethality and centrality in protein networks. Nature 411, 41–42.
[11] Kohn, K.W. (1999) Molecular interaction map of the mammalian cell cycle control and DNA repair systems. Mol. Biol. Cell. 10, 2703–2734.
[12] Simeonidis, E., Rison, S.C.G., Thornton, J.M., Bogle, I.D.L. and Papageorgiou, L.G. (2003) Analysis of metabolic networks using a pathway distance metric through linear programming. Metab. Eng. 5, 211–219.
[13] Fell, D.A. and Wagner, A. (2000) The small world of metabolism. Nat. Biotechnol. 18, 1121–1122.
[14] Brooke, A., Kendrick, D., Meeraus, A. and Raman, R. (1998) GAMS: A User's Guide, GAMS Development Corporation, Washington.
[15] Levine, A.J. (1992) Viruses, Scientific American Library, NY.
[16] Burgert, H.G., Ruzsics, Z., Obermeier, S., Hilgendorf, A., Windheim, M. and Elsing, A. (2002) Subversion of host defense mechanisms by adenoviruses. Curr. Top. Microbiol. 269, 273–318.
[17] Banks, L., Pim, D. and Thomas, M. (2003) Viruses and the 26S proteasome: Hacking into destruction. Trends Biochem. Sci. 28, 452–459.