Drug molecules and biology: network and systems aspects

Malcolm P. Young[1,2,*], Steven Zimmer[1] and Alan V. Whitmore[1,3]

[1]e-Therapeutics plc, Holland Park, Holland Drive, Newcastle, NE2 4LZ, UK
[2]Institute of Neuroscience, Newcastle University, NE2 4HH, UK
[3]School of Pharmacy, University of London, Brunswick Square, WC1N 1AX, UK

*To whom correspondence should be addressed

1 Biological robustness and therapeutic discovery

Therapeutic discovery is the process of finding ways beneficially to affect biology. The means by which therapy may be accomplished are very varied – small bioactive molecules, large ones, antibodies, vaccines, nucleic acids, radio-ligands, surgery and so on – but the task of finding new ways to affect biology beneficially with drugs has evidently proven challenging (Scannell et al 2010; Zimmer and Young 2009; Duyk 2003; Hellerstein 2008).

One reason for this challenge is that biology is (thankfully) extraordinarily robust to almost every kind of insult or assault, since evolution has selected biological function for this robustness (Kitano 2004, 2007; Wagner and Wright 2006; Wagner 2007; Masel and Siegel 2009; Masel and Trotter 2010; Raman and Wagner 2011). Whole organisms, and the important functional processes within them, have to be robust against variability in a wide range of parameters, including temperature, radiation, salinity, pH, oxygen partial pressure, loss or partial failure of their components, against invasion by parasites, and against changes in the chemical environment (Piersma and Van Gils 2011). To accomplish this, functional systems in biology are composed of a complex set of pathways, cycles, feedback loops, signalling sequences, homeostatic balancing mechanisms, redundancies, sub-networks and clusters, competitive interactions, long and short-distance interfacing systems and higher-order topological motifs (Young 1992; 1993; Scannell and Young 1993; Young et al 2000; Zimmer and Young 2009; Krakauer 2006). These network features intervene between adaptive, robust biological function, and the components of these networks, such as proteins, nucleic acids, lipids, second messengers and metabolites, which tend to be labile, transient, prone to damage and generally much less robust (Wagner 2007; Wagner & Wright 2007). An analogy might be that of military organisation, such as the organisation of soldiers into platoons, regiments and battalions, which groupings intervene between the individual soldier, who is prone to injury and incapacity during a battle, and a robust performance of the army.

Where biological systems malfunction, or where perfectly functioning biology is nonetheless undesirable for humans (for example in disease caused by pathogens), therapeutic discovery has to confront the complexity and robustness of biological systems head on (Hellerstein 2008; Hopkins 2008; Hase et al. 2009). Success in doing so results in efficacious and safe new medicines (Hopkins 2008; Zimmer and Young 2009). Failure to do so results in weak efficacy, development of therapeutic resistance, poor side effect and tolerance profiles, drug candidate failures in the clinic more often than not, and unsustainably low productivity (e.g. Scannell and Blanckley 2010).

The best way to approach therapeutic discovery is more in dispute now than at any time in the last many decades, not least through a slow but increasingly general acknowledgement of the astonishingly poor productivity of conventional discovery programmes (Duyk 2003; Kola and Landis 2004; Hopkins 2008; Kola 2008; Scannell et al 2011), and the slow realisation of the ramifications of advances in network biology and network mathematics for discovery (Csermely et al 2005; Hopkins 2008; Pujol et al 2010; Young and Whitmore 2010; Zimmer and Young 2009).

To gain increased traction in discovery, do we need to consider and address network properties rather than the properties of single components, such as protein "targets"? Is inhibition of one protein "target" ever a scientifically plausible explanation of therapeutic function, or indeed a scientifically plausible motivation for systematic drug discovery? Is one "target" ever enough?

2 Biological networks and their properties

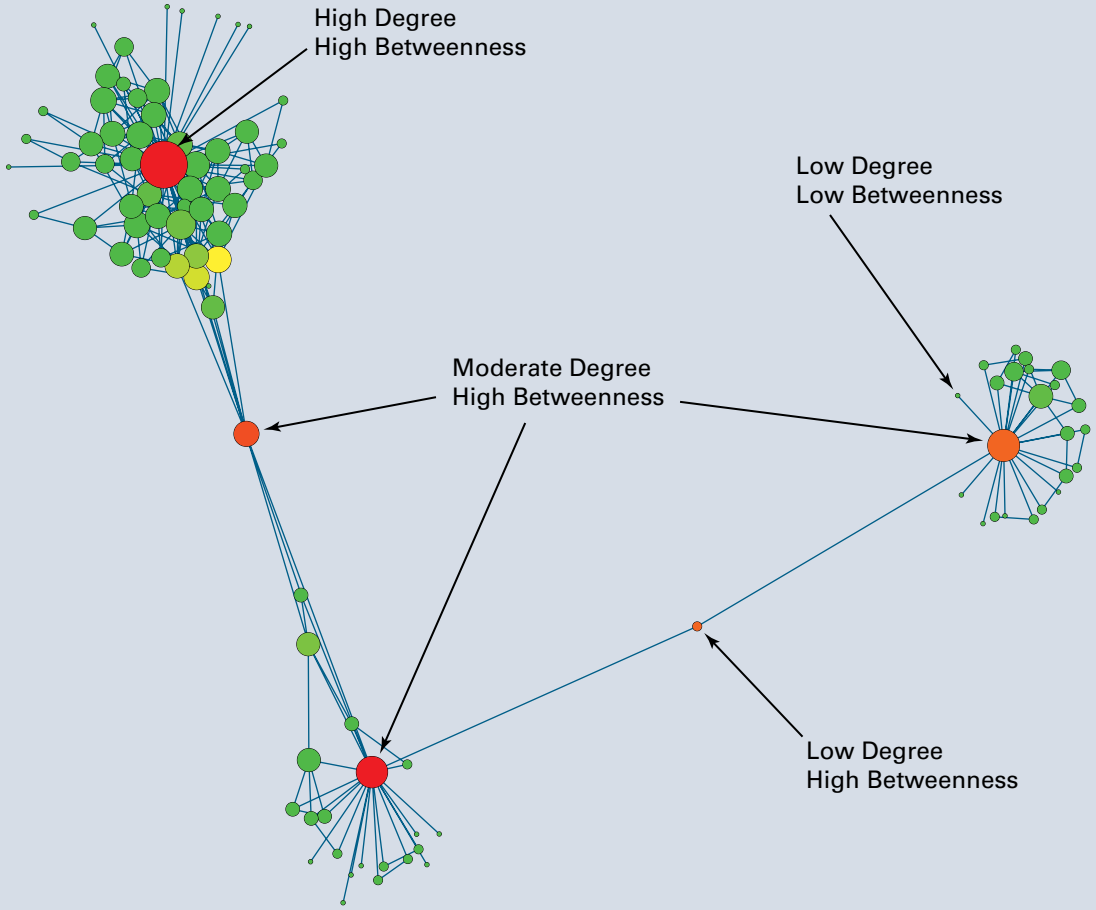
An encouraging development in network science has been the discovery that networks of biological complexity are strikingly robust to deletion or interruption of their components (e.g. Young et al 2000; Albert et al 2000; Barabasi and Oltvai 2004; Albert 2005). These results are encouraging because they are in keeping with the expectation that it is these network structures and topologies that bridge between unreliable components and reliable systems-level function in biological systems, and so they reduce the mystery of how biology is able to generate such robustness from such unreliable components (Wagner 2000; Pujol et al 2010).

A further interesting feature of the analysis of biological networks is that individual components of them (nodes) differ greatly in importance for the network. For example, in many biological networks, "degree", - the extent to which a node is connected to other nodes - varies in a way that approximates a power law, so that a very small number of "hubs" have an unusually great number of connections (Barabasi and Albert 1999; Barabasi and Oltvai 2004). This network property has been related in several datasets to the biological essentiality of network components; and to impact on network integrity, the greater the degree of the node to some extent reflecting the greater impact on the network from its deletion (e.g. Young et al 2000; Jeong et al 2001; Kaiser et al 2007).

The gross connectedness of nodes, though, is only one of many ways in which nodes can have differential significance in a network of biological complexity. Nodes that are entirely unremarkable in terms of degree can nonetheless be highly significant in the structure of the network because they bear other structurally important properties. These other structurally important properties are numerous, and include, for example, "betweenness", the extent to which a node connects otherwise largely unconnected clusters, and "redundancy", the extent there are pathways through a network other than through a specific node (Paolini et al. 2006; Kaiser & Hilgetag 2004; Young et al 2002a; 2002b). In addition, it is known that nodes with high values in these properties are associated statistically with biological essentiality in well-studied systems (e.g. Idowu et al 2004, 2005; Hopkins 2008). Hence, network components can vary greatly in their significance for network structure and function; this significance can derive from rather many different network properties; and across many of these network properties, there is evidence of statistical association with biological importance (e.g. Idowu et al 2004; Jeong etal 2001). Biological network structure is illustrated in Figure 1, which exemplifies some of the network properties that have proven informative.

Figure 1: Measures of node significance in a network

An example of a sub-component of a real biological network (protein-protein interactions around the microtubule associated protein Tau). A number of network properties referred to in the main text are illustrated here. The degree of each protein node (the number of neighbours to which it is connected) is represented by the size of the circles (low degree-small). There are a small number of nodes with high degree and many more having low or intermediate values. The parameter betweenness (see text and glossary) is also represented on a colour continuum from green (low values) through yellow (intermediate values) to red (high values). One or two nodes with high degree also have high betweenness but a significant number have only low or moderate degree and are nevertheless critical for connecting subclusters of this network; consequently they have high betweenness and appear bright red. Examples are pointed out by the black arrows on the diagram.

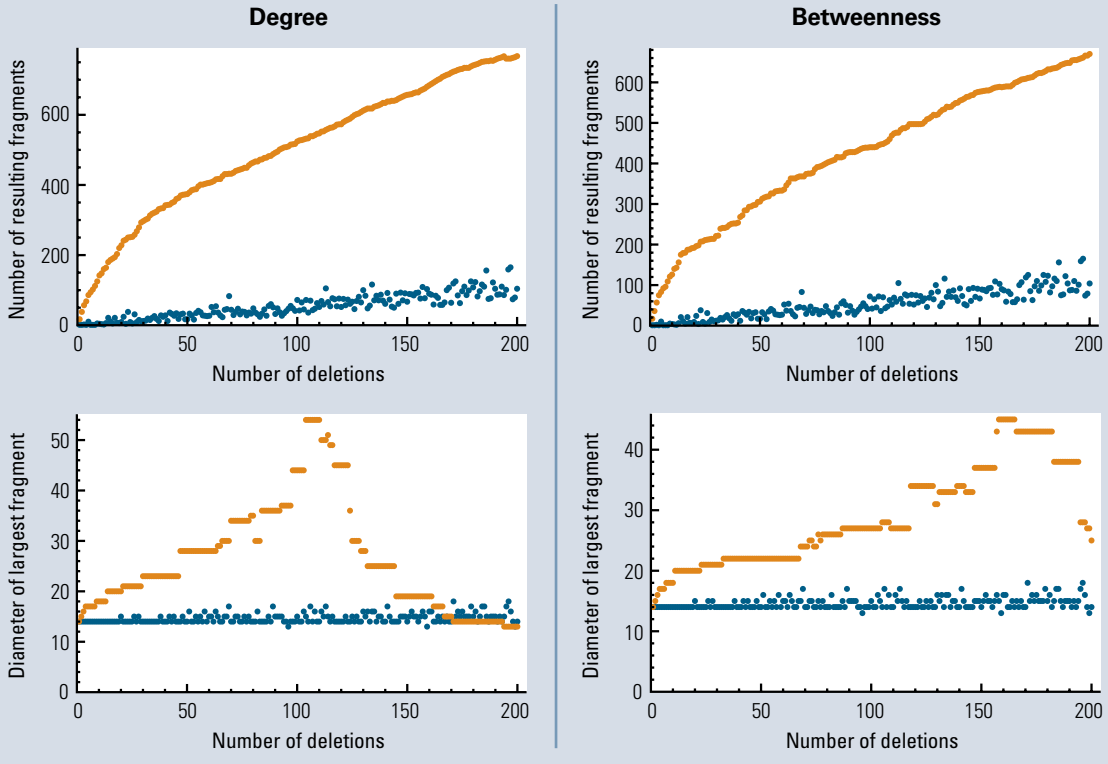


The association of these network properties with biologically important features suggests that networks should be susceptible to attack or intervention that is guided by knowledge of these network properties of nodes. And indeed this is what has been observed in "deletion experiments", in which the effects on a measure of network integrity are recorded when cumulative deletions are made in the network (e.g. Young et al 2000; Jeong et al 2001; Kaiser et al 2007). Comparisons between cumulative deletions prioritised by several different network properties with deletions prioritised at random show that networks are highly robust to random deletions, but much more susceptible to targeted ones (see Figure 2). This result is helpful for the prospects of systematic drug discovery, since it suggests ways of prioritising approaches to therapeutic intervention, at least to the extent of identifying "targets" in a scientifically rational way (Zimmer and Young 2009).

These results, though, offer a direct scientific challenge to the central precept of conventional single-target drug discovery (STDD) activities. For STDD to be a plausible approach to discovering new medicines, a single targeted intervention should be sufficient to affect biological network integrity significantly. However, to date, no known network property, no plausible measure of the impact of an intervention on network integrity, and no biological network that has been analysed has shown that a single deletion can deliver significant change to network integrity and so function (Wagner, 2000; Agoston et al, 2005; Hellerstein,2008; Hopkins, 2008). This is, in one way, not at all a surprising result: returning to our military analogy, there is no single soldier we can shoot whose demise would significantly affect the performance of the army. Even shooting the General would only affect the army significantly in the very special circumstance that his functions cannot be replaced by his general staff (Young et al 2002a; 2002b). These results are illustrated in Figure 2, which shows the effects of one and many deletions prioritised by a variety of different network properties on two commonly used measures of network integrity.

Figure 2: Multiple targeted interventions are required to perturb biological networks

The effects on integrity of a network of biological complexity (the giant component of the yeast interactome downloaded from http://interactome.dfci.harvard.edu/S_cerevisiae/) when progressively increasing numbers of random (blue squares) or targeted (orange squares) node deletions are made. Two measures of node importance were used for targeting: degree and betweenness, as illustrated in Figure 1 and defined in the main text and glossary). Two measures of network integrity are illustrated for each type of targeted intervention – the number of disconnected sub-networks into which the network has been broken by the interventions (upper panel), and diameter, which is the longest shortest path between any pairs of nodes in the network (lower panel). Deletions prioritized by such measures of significance have a much greater impact than random ones (which have almost no effect until a very large proportion of nodes have been removed). Even with prioritization it is still necessary to make multiple interventions to perturb the network significantly.



With only one known important exception (see below), multiple targeted interventions in biological networks always produce greater network impact than a single intervention. Viewed from this perspective, a strategy for developing drug candidates with modest or insufficient efficacy would be to base one's discovery efforts on STDD. Conversely, a reasonable framework for developing efficacious ones would be to try to target multiple network vulnerabilities in disease-related cells, informed by network analysis, and hence to follow a multi-target drug discovery (MTDD) strategy.

3 Multiple interventions: expect the unexpected

There is one known and potentially very important exception to the observation that multiple interventions always yield greater network impact than single interventions. This exception relates to the possibility of negative synergies between multiple interventions, in which one intervention negates or diminishes the effect of another. These effects are "paradoxical" in that function pre-dating the interventions may be restored by further interventions in a system affected by the first intervention. For example, in disease, the original dysfunctional state may be inadvertently restored if one intervention, intended to improve the condition, is accompanied by another intervention made in an unfortunate place in the network. Such paradoxical effects are well known and characterised in, for example, neurology where they are known as "paradoxical restorations of function" (e.g. Sprague 1966; Payne et al 1996; Young et al 2000; Hilgetag et al 1999). There is ample opportunity for these effects to manifest in any biological network where there is both sub-network structure and competitive interaction between sub-networks, as appears to be the case in all biological networks of relevance to drug discovery (e.g. Hilgetag et al 1999; Young et al 2000; Spirin and Mirny 2003).

The simplest case in which these negative synergetic effects manifest occurs when two simultaneous interventions in the network are made (Young et al 2000). In this case, one intervention might negate or diminish the other, by, for example, restoring balanced competition between competing subsystems when such balanced competition is an important feature of system function (e.g. Hilgetag et al 1999; Young et al 2000). In the context of drug discovery, these effects might manifest as lack of experimental efficacy despite one's drug candidate certainly diminishing the function of an important "target" protein – because, unbeknownst to the drug designer, the molecule also intervenes on a protein in a different sub-network that has competitive interactions with the sub-network containing the primary "target", restoring function based on balanced competition. Note that studying only the local pathway in which the primary target is believed to participate would not be sufficient to anticipate these paradoxical synergies, which are network-based, systems-level effects. Note also that without checking for these types of synergetic effects, efficacy is not reliably predictable from potency at a single "target" in the general case (Young 2008; Zimmer and Young 2009).

If the presence of a bioactive molecule in or near a cell involves any pleiotropy (effects on proteins or other signalling molecules downstream in the network from a directly affected protein), promiscuity (direct effects on more proteins than the designed "target"), or polypharmacy (more species of molecule present than the bioactive molecule itself – such as the molecule's long-lived metabolites, each of which can exhibit both pleiotropy and promiscuity), then multiple interventions in the network of higher order than two will be present. We examine whether and how drug molecules intervene on multiple proteins in the next section. However, determining the net effect of more complex sets of interventions in biological networks than two is a combinatorial problem that is very difficult to address

experimentally, since large numbers of permutations cannot effectively be explored in that way. The need to determine the net effects of these synergies as early as possible in drug development, in order to de-risk candidates prior to expensive clinical trials, means that there is an irreducible computational step in discovery and development programmes that do not wish to rely on luck (Young et al 2002a; 2002b; Young 2008; Zimmer and Young 2009).

4 Effective drugs and multiple interventions in networks

The likely presence of negative synergetic effects from making multiple interventions in networks of biological complexity might suggest that MTDD is a riskier strategy than STDD, since intervening at only a single "target" would not engage these factors. This would only be true, however, if drug molecules motivated by STDD did in fact only intervene at the designed "target". Hence, an important question is: do any drugs affect only one protein "target"?

We noted above some theoretical reasons for doubting that any efficacious drug only affects a single "target". Chemical biology also suggests that the assertion that efficacy derives from a single "target" intervention is a very strong scientific claim, perhaps one so dubious as to require special evidential support (e.g. Hopkins 2008; Young 2008). A molecule designed to have very high affinity at one protein is likely also to have affinities at other proteins, since proteins are configurationally similar. This promiscuity has been discussed extensively recently (Hopkins et al 2006; Morphy et al 2004; Paolini et al 2006; Roth et al 2004; Weitz-Schmidt 2001), and is illustrated by the spectra of affinities shown across screens for binding affinity with other proteins (Hopkins et al 2006). The binding spectra so far explored in the literature relate only to a very small proportion of the protein *dramatis personae* in human and pathogenic cells, typically being fewer than 100 proteins (e.g. Hopkins et al 2006). However, while the number of proteins in human and other disease-related cells is not known definitively, it is likely that it exceeds many hundreds of thousands (Jensen 2004). Hence, there is great scope for drug-protein promiscuity in medicine, only a small proportion of which is typically addressed in discovery programmes. It cannot safely be assumed that these off-target interactions will not have consequences for efficacy (see Section 3) or safety, and we hypothesise that these unpredicted effects are one reason for the prevalence of clinical stage attrition of drugs motivated by STDD, since clinical trials test expensively for the global effect of the presence of the molecule in patients, including these currently poorly attended off-target interventions (Young 2008).

The presence of a molecule in the body can affect the functional contribution of proteins for which the molecule does not itself have high affinity. For example, the presence of a molecule can pleiotropically affect the expression and abundance of proteins to which it does not bind directly. A celebrated example of this is the effect of CGK1026 on the expression of telomerase (Won et al 2004). CGK1026 forms a complex with HDAC2 and E2F, which has the effect of de-repressing TERT (Won et al 2004). Hence, CGK1026 profoundly affects the expression of telomerase, with potentially significant effects on cell biology, without binding telomerase directly. Similarly, the presence of a molecule can pleiotropically affect the phosphorylation or other functional state of proteins to which it does not bind (e.g. Liao and Laufs 2005; Greenwood et al 2006). Indeed, it can be argued that the importance of kinases as therapeutic "targets" derives largely from the pleiotropic effects they exert on many other proteins (Morphy 2010). There is evidence for even less direct pleiotropic effects in networks of biological complexity: deletion of nodes can readily affect other unusually significant nodes through network-mediated effects via these nodes' network associates (Won et al 2004; Young et al 2000; Zimmer and Young 2009).

Binding promiscuity and these several varieties of pleiotropy can also attend each downstream metabolite of a bioactive molecule that is sufficiently long-lived to interact biologically. Hence, the presence of a molecule in the body can entail a pattern of interventions in networks far beyond its designed primary "target". These wide-ranging effects can be mediated through wider promiscuous binding; through the pleiotropic downstream effects of each of these interactions; and its metabolites, if sufficiently long-lived to be active, can each also entail wide-ranging effects mediated through their own promiscuity and pleiotropy. Hence, most, and perhaps all, bioactive molecules cause multiple interventions in networks of biological complexity, and so will engage the potentially complex consequences explored above (Young 2008; Zimmer and Young 2009).

In regard to defining the best way - post the recent advances in network science - to approach therapeutic discovery, drugs motivated by STDD principles are hence no better placed to avoid the problem of the paradoxical effects of multiple interventions in networks, since these drugs are almost certain themselves to be multiple interveners. Motivating discovery by MTDD principles is better supported theoretically, and also perhaps psychologically, since MTDD programmes might be disposed to be more alive to a potentially wide range of interventions caused by their candidate molecules. In the cases of both STDD and MTDD strategies, though, network analysis suggests that there may be some further hills to climb before drug discovery and development will generate noticeably higher productivity.

5 Drug discovery problems in light of network science

Considering impact graphs, like those in Figure 2, suggests that greater impact attends quite large numbers of network interventions. If network impact is a good proxy for efficacy, then, in general, the more "multi" a multiply intervening drug is, the more effective it should be, all else being equal. Furthermore, it has been suggested, on the basis of research on the transcriptional regulatory networks of bacteria and yeast, that such interventions can be quite weak and yet remain very effective in combination (Agoston et al 2005). This implies that it would be reasonable to try to design for many more than two "targets" (Morphy et al. 2004, Wermuth, C. 2004, Hopkins et al, 2006; Metz, J.T and Hajduk, P.J 2010). The technical difficulty of doing this is addressed elsewhere in this volume. However, we have seen that there are potentially important negative synergies that can be engaged should some of our molecule's interventions be in unfortunate places in the network, the effect of which could be to greatly diminish efficacy, even when a number of interventions in biological networks are made. We have also seen that drug molecules tend to make rather many interventions in biological networks, in a number of different ways.

These considerations suggest some scientific elements that may underlie low productivity in drug discovery. Official statistics suggest that only around 1 in 10,000 lead-optimised molecules eventually becomes a marketed drug (European Commission, 2008). Further observations are that discovery and development across the industry have turned exponentially improving inputs (e.g. bases sequenced per hour; compounds synthesised and screened per hour) into exponentially decreasing output (e.g. new drugs gaining market approval divided by R&D spend) over many decades; and that the Net Present Value of small molecule-oriented R&D is now negative when considered across the industry as a whole (Scannell and Blanckley 2010; Scannell et al 2011). Before becoming overly enthused by new approaches to drug discovery, it would perhaps be prudent to consider some scientific explanation of what has gone wrong with the conventional approach, so that the same or similar difficulties can be avoided in the future.

A consequence of the features of promiscuity, pleiotropy and metabolite-polypharmacy (Section 4) may be that the great majority of the interventions in biological networks that are caused by the presence of a drug or candidate are not the one (or two) that motivated its design. Even when a molecule binds at high affinity to its designed "target", it may affect many other biological system components in an undesigned way. Since unfortunate undesigned interventions can diminish both efficacy and safety, and fortunate undesigned interventions can occasionally improve both efficacy and safety, a poorly controlled factor would appear to be present in drug discovery programmes that do not take account of these features. From first principles, we expect that this uncontrolled factor would manifest as great difficulty in reliably predicting efficacy and safety in patients from preclinical information. This does indeed seem to be a feature of current problems in drug discovery and development (Duyk, 2003; Hopkins 2008; European Commission 2008; Young 2008; Zimmer and Young 2009).

Very many more candidate drug molecules are developed than go on to become effective drugs (European Commission 2008), with very many failing for efficacy or safety reasons in clinical evaluation with human volunteers or patients. The large number of failures and the small number of successes provides a very strong opportunity for *selection* to operate. Observations suggest that across very many discovery programmes, there is an occasional success, usually offering an incremental benefit in efficacy or safety, framed by many candidates failing in the clinic (Scannell and Blanckley 2010). It is tempting to believe that the successes are derived by rational design. Confidence in rational design might be reinforced by survivor bias, that is, podium presentations being biased toward projects that delivered clinical results broadly those being looked for, and biased against those that did not (Scannell and Blanckley 2010). However, an equally plausible explanation, taking account of the large number of failures in "targeted" development programmes, is that the successes are the result of *clinical selection* operating over plausible chemical diversity, and not principally the result of design. Each candidate will doubtless possess high affinity for its designed target, and it will also likely cause a variety of undesigned interventions that mean that its efficacy and safety proposition in human beings will be highly unpredictable from its affinity for its primary "target".

Clinical evaluation then tests for the global net effect on efficacy and safety of the presence of the molecule in patients, including all the "off-target" interventions, arising from promiscuity, pleiotropy and metabolite-polypharmacy. Most combinations of designed and undesigned interventions would be expected to be useless or unsafe (and often both), since biological networks are robust to extensive random interruptions. Clinical evaluation will fail these candidates very often, at great cumulative expense. But occasionally the combination of designed and undesigned interventions would be expected to be beneficial, simply by chance. The net effect, evaluated by the clinic, would support these rare candidates, and they have opportunity to proceed through development to become successful medicines.

This model of clinical selection over plausible chemical diversity - as opposed to medicinal chemistry design - suffers from the great expense of clinical trials, and the great number of them that are required for even a small number of successful new drugs to emerge (di Masi et al 2003; Kola and Landis 2004). The model is sustainable only when the revenues from the occasional success are sufficient to support the large number of clinical failures each success requires. If the bar is raised in any way, however, through, for example, it being more difficult to charge a premium for new drugs that are only marginally superior, or more aggressive generic erosion, or increased regulatory distaste for safety risk, or areas of remaining unmet medical need being harder to gain traction in using conventional methods, then it will generally fail, no matter what the scale of resources that is dedicated to it.

Even if it remains difficult realistically to design molecules with specific multi-intervention profiles, there may be ways to increase the efficiency of a selection-based model of drug development. One way would be to pay much greater attention to the potentially wide-ranging interventions caused by candidate drug molecules, to pay much greater attention to the network impacts of these interventions, and to reject a far greater proportion of drug candidates on this basis before they reach the clinic.

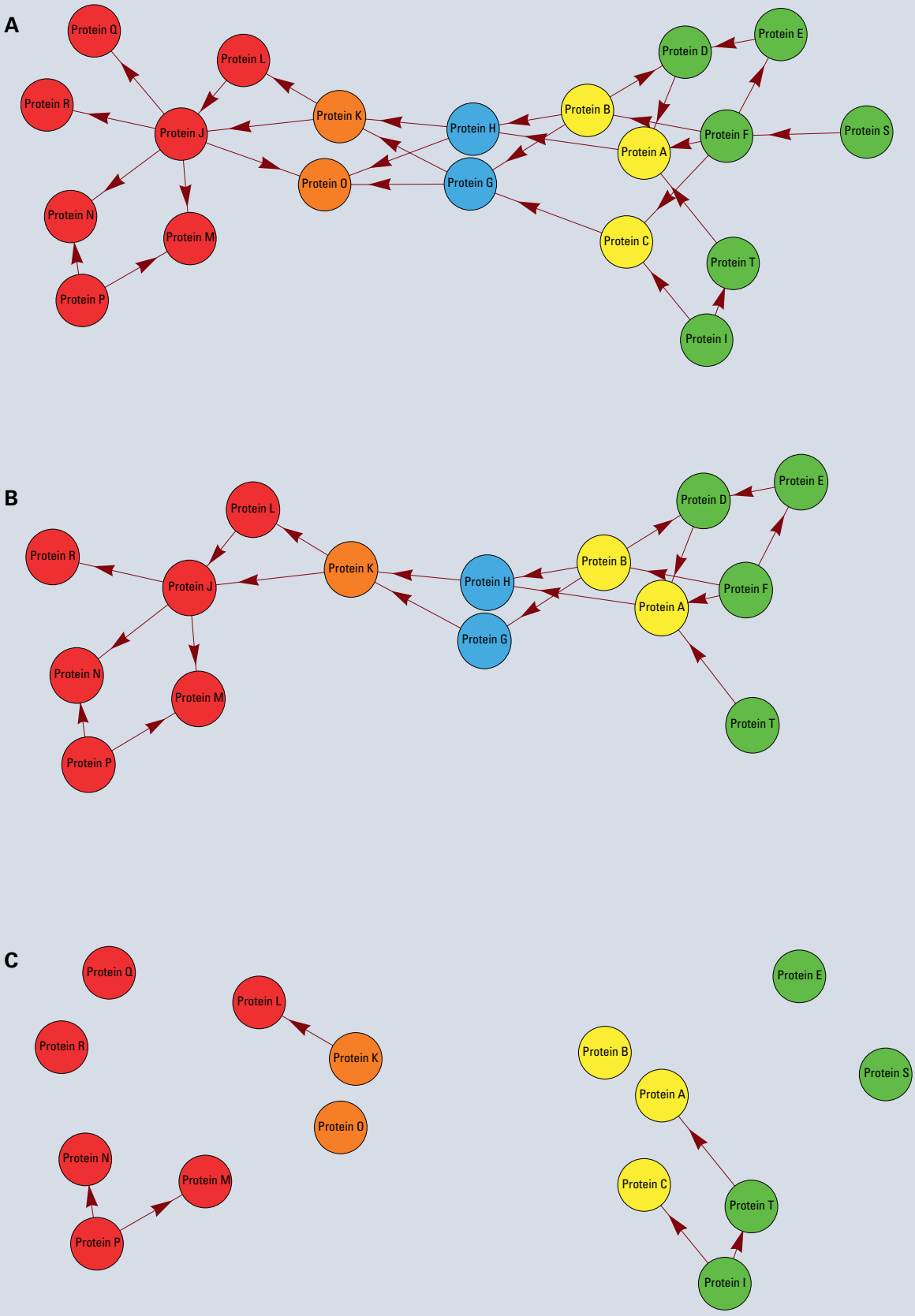
6 Network Pharmacology: exploiting advances in chemical biology and network science

In what ways can improved understanding of chemical biology and network science be leveraged to improve upon the unsustainably low productivity of conventional, typically single "target", drug discovery? Historically, heuristics that were insufficiently productive to be sustainable have often been replaced by approaches based on the application of engineering principles (e.g. Barlow 1998), and this is the intent of Network Pharmacology (Young 2008; Zimmer and Young 2009). In the case of improving the productivity of drug discovery and development, the pressing needs are to obviate failure during the process, and to deliver high efficacy in areas of unmet medical need, safely.

Since all bioactive molecules intervene on multiple proteins (see Sections 3 and 4 above), we consider particularly the case of the network impact of multiple interventions in complex biological networks. There are very many different combinations of, for example, five proteins that can be removed simultaneously from any network of full biological complexity without affecting network integrity in any significant way. This reflects evolutionary selection in favour of biological systems that are robust to random damage and chemical perturbation, through redundancy, amongst other organisational adaptations. But a small number of combinations of five proteins can have a very great impact on network integrity in most biological networks. The distributions of network impact for all metrics so far tested for a given number of interventions tend to be exponential, showing a long tail of irrelevant combinations, with a small number of highly impactive ones (see Figure 3).

Combinations of simultaneous interventions that yield very high network impact represent only a tiny proportion of all the possible combinations of interventions. Such high-impact combinations are unlikely to be found unless they are specifically optimised or searched for. These combinations yield many-fold greater impact than those achieved through single designed interventions; and from single designed interventions accompanied by a number of undesigned interventions. From first principles, therefore, where network impact is a good predictor of efficacy in target cells, much greater efficacy would be expected from optimizing for network impact through selecting specific combinations of interventions, than could be derived from intervening on one (or two) "target" proteins, with or without associated undesigned interventions. Bearing in mind that interventions at specific proteins can be weak in terms of binding affinity, and indeed pleiotropic, yet be highly effective in combination (e.g. Agoston et al 2005), greatly enhanced efficacy should be obtainable by optimizing for network impact, rather than nanomolar binding potency at a single protein "target". From this perspective, network pharmacology optimizes a completely different feature of chemical biology than conventional discovery, and so represents a significant, scientifically motivated, departure from conventional approaches. Similarly, where conventional discovery processes are industrialized around optimizing for nanomolar potency at single "targets", as they appear often to be, this may in large part explain the exponential decline in conventional productivity as the outcome of industrialising design for a largely irrelevant feature (see also Swinney and Anthony 2011).

Figure 3: Multiple interventions can be optimized by combinatorial impact analysis



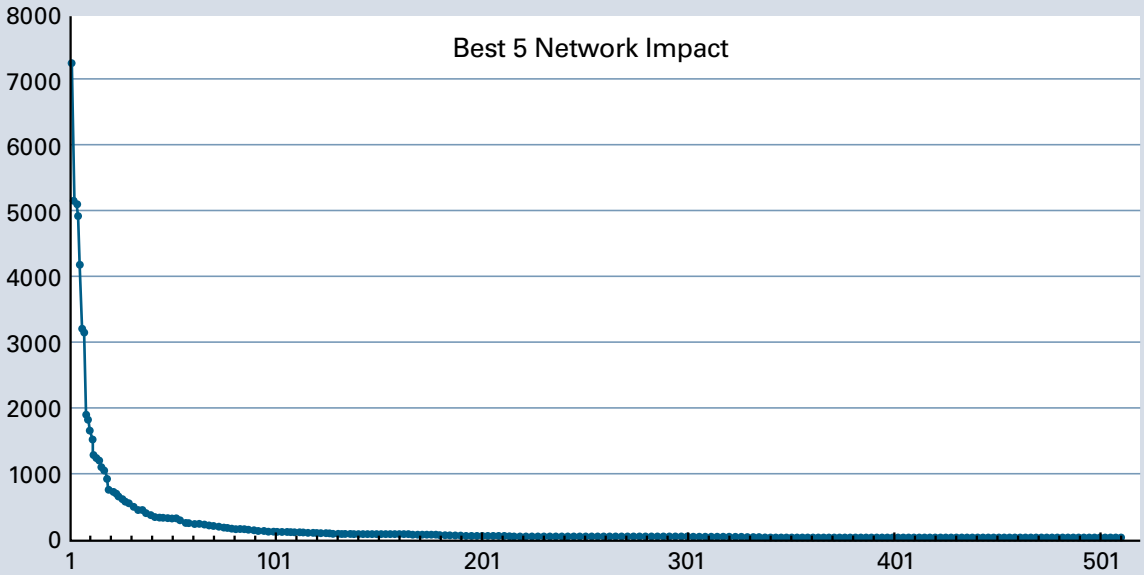
In **A** a simplified artificial network with realistic properties is represented. It controls the systems level function 'Build a Cell Wall" for a hypothetical pathogenic bacterium. The objective is to find a way to disrupt it maximally using a maximum of 5 interventions. Green circles are proteins responsible for post-translational modification of newly synthesized cell wall components. The cell wall components are shown in yellow. The cell wall proteins must then interact with the carrier proteins shown in blue to form a complex (shown in orange) that can be assembled by the cluster of wall assembly proteins (in red). For a viable systems-level function to be present there must be a continuous path available between representatives of each type (colour). The network shows considerable redundancy – there are a number of alternative possible paths from green to red.

In **B** five nodes have been deleted at random. This has had no material effect on the ability of the bacterium to assemble a cell wall: there remains a viable path from green to red.

In **C** five targeted deletions have been made: the highest degree nodes have been removed; and the two nodes with highest betweenness have also been deleted. The redundancies have been overcome and the network has disintegrated - the bacterium can no longer make a cell wall. In fact in this example it might have been sufficient just to delete the two blue nodes to disconnect the 'parts department' from the 'assembly line', although some components might still have reached the assembly complex by simple diffusion if this were not also disabled.

In **D** we show an example distribution of combinatorial impacts on network integrity for a real biological network. Following on from the example above five nodes are removed at random and the impact of the deletion calculated for the whole network. Most ways of removing five nodes yield a very minor impact on network integrity. However, there are a small number of combinations of five node deletions that have a profound effect on network integrity. All single node deletions would be placed far to the right in the distribution, yielding negligible impact on network integrity, and most combinations of a designed single deletion with four other deletions associated with the single deletion semi-randomly would have negligible effect on network integrity. Hence, the strategy of network pharmacology drug discovery is to find the very few highly impactive sets of proteins that should be addressed simultaneously, and to develop molecules that deliver these optimally impactive sets of interactions, in order to yield high efficacy.

D



Employing network pharmacology in drug discovery requires the optimisation step of combinatorial network impact analysis to be embedded in a number of other processes. Network pharmacology is relatively new, but most discovery projects motivated in this way have begun by identifying a specific medical need that is poorly served by existing therapies, which would represent a sufficient market to justify the expense of development, and which can be framed in an affordable and implementable development and regulatory pathway. Once the specific medical need is identified, the pathophysiology of the disease is explored; the optimal medical intervention point in the pathophysiological development of the disease determined; and systems-level functions operating at this point in the pathophysiology mapped in detail. The genomic and proteomic networks that are implicated in these functions are then curated, and then combinatorial impact analysis is undertaken to derive optimally impactive sets of proteins to intervene upon simultaneously in order to drive the network(s) in a therapeutically desirable direction.

Once optimally impactive sets of proteins have been derived for the disease, molecules are sought whose presence in the human body is associated with an appropriate pattern of interaction with proteins in the optimal set. These interactions can be exerted through direct, allosteric and promiscuous binding affinity; or pleiotropically, through effects on protein expression, phosphorylation, or post-translational modification; or through indirect network-mediated effects, typically by affecting a near topological neighbour of a protein in the optimally impactive set. This process requires significant chemoinformatics resources, but the inclusion of non-binding interventions in the mapping between an optimally impactive protein set and molecules that could deliver it has the effect of opening the bandwidth of molecular selection beyond that available from direct high-affinity binding alone. Molecules that deliver an optimally impactive set of interventions are likely to be efficacious in delivering the specific therapeutic functionality sought in the design of the discovery project.

The principle of combinatorial network impact, employed to focus discovery on high efficacy above, can also be applied very early in development to de-risk candidate molecules for safety and tolerance. Conventional *in vitro* and *in vivo* preclinical toxicological studies plainly do not de-risk molecules for safety in human patients nearly well enough, evidenced by the high number of safety failures in clinical stages (e.g. DiMasi 2001; Kola and Landis, 2004). However, insight into the effect on a wide variety of normal human cell-types of the presence of a candidate molecule in the body can be derived by calculating the combinatorial impact of those proteins affected by the molecule in each of hundreds of normal human cell-types. These cell-types represent as far as current bioinformatics data are able, the full range of tissue types. This process requires significant chemo- and bioinformatics resources, and, unlike targeting for efficacy, is undesirably open-ended as a scientific problem, since there are very many different cell-types in patients, and data on the networks within them are far from complete. Notwithstanding the data limitations, which are lessening as knowledge of normal function increases, it is certainly feasible to triage a proportion of molecules with the potential for significant impact in these normal cell-types; and these can then be discarded.

This workflow embeds the process of combinatorial network impact analysis and the promiscuity and pleiotropy of bioactive molecules to derive molecules that are probably both efficacious in the specific indication sought, and safe. The workflow is designed to predict accurately what risks a candidate has as cheaply and early as possible, and does not rely on expensive clinical trials as the principal means of de-risking molecules. In terms of the effect of this approach on discovery and development productivity, conventional processes cannot predict either clinical efficacy or safety accurately at present, and so most drug candidates fail after substantial sums are spent on them, since expensive tests and trials are the only

way to evaluate these risks (e.g. Adams and Brantner, 2006). This means that at any point the vast majority of discovery and development resource is tied up in drug candidates that are going to fail, and productivity is unsustainably low. But if early stage predictions could be improved, a lower proportion of candidates that are predicted to be worth developing will fail, and expensive trials will usually evaluate the risks as low. In this scenario, at any one time, the majority of discovery and development resource will be invested in drug candidates that will probably not fail because technical risks manifest, and productivity will be much higher. Commercial risk for individual companies would not diminish in this scenario. Indeed it will rise if more good drugs come to the market, but that would perhaps be a better problem to have, especially from the perspective of patients.

7 Prospects for multi-target drug discovery in light of network science

The advances in network science we've touched on in this chapter hold out great promise for enabling systematic multi-target drug discovery into the future.

- First, the robustness of biological systems, which confronts head-on all attempts to change biological systems for the therapeutic better, may be mediated by network features that we can now target specifically, through network analysis.
- Second, network science advances throw a harshly critical light on some of the fundamental principles of most drug discovery activity, such as the single "target" idea, suggesting that resources could be diverted toward activities more likely to be productive, such as MTDD strategies. There may be a need to influence the education of scientists in parallel with any paradigm shift in the industry.
- Third, some of the pitfalls of making multiple interventions in networks of biological complexity, such as paradoxical synergies between interventions, are becoming clear in a form that permits empirical testing and navigation around them.
- Fourth, the exponential distribution of combinatorial impact in networks, together with the mediocre positioning of both single "target" and poorly targeted multi-"target" drugs in these distributions, suggest that much greater efficacy than seen hitherto could be engineered into drug molecules using this type of design principle.
- Fifth, the many different ways in which bioactive molecules can cause interventions in biological networks are becoming clearer, enabling the prospect first of rational selection of molecules from their chemoproteomic signatures for specific medical tasks, and, perhaps in future, the design of molecules on these principles.
- Sixth, diseases that have proven highly resistant to conventional therapeutic discovery, which diseases tend to be those in which a new drug can still be priced at a premium, are also those that seem unusually suited to therapeutic targeting by approaches based on these principles of network pharmacology.

Glossary

Betweenness	A network property of nodes (or their connections) in a network. It quantifies the frequency of appearance of an individual node (or connection) in the set of all shortest paths within a network; a measure tied to the importance of such a node or connection in establishing information flow through the network
Cluster or sub-network	Community structure within a complex network: in the simplest case a group of nodes that share more connections with each other than with other nodes in the network; or in functional terms an interconnected group of nodes whose interactions underlie an identifiable, emergent, systems-level function.
Combinatorial impact	Denotes the impact on the structural integrity of a complex network of removing an optimal set of nodes and/or connections.
Degree	The number of connections that a node has with other nodes in a network.
Essentiality	In biological networks, this denotes the importance of a node (such as a gene or protein) to the cell's viability or to the ability to sustain a systems-level function.
Heuristic	An experience-based strategy or tactic that, using available information, is intended to yield the desired outcome.
Hub	A highly connected network node; a node that has high degree.
Impact (network)	A measure of the importance of an intervention on the integrity or functional state of a complex network. Metrics include the histogram of the shortest path; or a plot of the magnitude of the longest (or average) shortest path in a network versus different types of intervention(s).
Integrity (network)	The structural state of a network in which the network is able to function normally to generate its emergent, systems-level function.
Metabolite Polypharmacy	In this context is used to represent the situation where not only a compound but also its (possibly multiple) metabolites are pharmacologically active and therefore have additional (and often unpredicted) network consequences.
Negative Synergy	An interaction between two or more interventions within a complex network whose net impact on the emergent function of that network process is either neutral (because they cancel each other out) or has an effect opposite to that intended.

Network Analysis	Denotes the analysis of the structure and nature of relationships between nodes and edges in a network.
Pleiotropy	In this context describes potentially multiple consequences immediately downstream of the effect of a bioactive molecule at a particular protein, which may change the phosphorylation state, abundance, methylation, post-translational modification of multiple other proteins.
Promiscuity	Denotes the ability of a molecule to interact with many different proteins either through the 'active sites' of those proteins or through allosteric binding.
Redundancy	This denotes a mechanism through which complex networks generate robustness to loss of components or to other insults. The degree of fault tolerance in these networks is related to the presence of network redundancies (e.g. alternative information flow pathways; or clusters with redundant sub-functions). In biology, redundancy is a fail-safe mechanism for systems-level networks permitting organisms to survive mutations or environmental insults with minimal or no consequence to the systems-level function controlled by the network.
Robustness	Usually refers to the ability of a complex system or network to maintain its functional characteristics in the face of perturbations. In biology, it can mean the ability of an organism to maintain a particular systems-level function or phenotype in the presence of mutations or external stressors. In network science, it denotes the ability of a network to withstand attack or component failures.

References

Adams C, Brantner V. Health Aff (Millwood) 25 (2): 420–8 (2006).

Agoston V., Csermely P. and Pongor S., Physical Review E 71, 051909 (2005)

Albert R., Jeong, H. and Barabasi, A.L, Nature; 406: 378-382 (2000)

Albert R. J. Cell Sci 118, 4947-4957 (2005)

Barabasi A-L. and Albert R. Science, 286, 509-512 (1999)

Barabasi A-L. and Oltvai Z. Nat Rev Genet 5, 101-113 (2004)

Barlow R.E. Engineering Reliability. ASA and SIAM, New York (1998)

Csermely, P., Agoston, V. and Pongor S. TiPS,26: 178-182 (2005)

Di Masi JA Clin Pharmacol Therap 69: 297-307 (2001)

Di Masi JA Hansen RW and Grabowski HG J. Health Economics 22: 151-185 (2003)

Duyk G. Science 302, 603-605 (2003)

European Commission, Innovative Medicines Initiative: Better tools for better medicines, page 5, Luxembourg: Office for Official Publications of the European Communities, 2008.

Greenwood J, Steinman L and Zamvil S Nature Rev Immunol 6, 358-370 (2006)

Hase, T., Hiroshi, T., Suzuki, Y., Nakagawa, S. and Kitano, H., PloS Comp.Biol.,5: 1-9 (2009)

Hellerstein M J. Pharm. Exp. Ther., 325: 1-9 (2008)

Hilgetag CC Kotter R and Young MP. Progress in Brain Research, 121:121-141 (1999)

Hopkins A. Nature Chemical Biology 4, 682-690 (2008)

Hopkins A.L., Mason, J., Overington, J.P., Curr. Opin. Struct. Biol., 16:127-136 (2006)

Idowu O. and Andras P. Proceedings CIMED (2005)

Idowu O., Lynden S., Young M. and Andras P. Proceedings IEEE Computational Systems Bioinformatics Conference Stanford, USA, California 623-625 (2004)

Jeong H., Mason S., Barabasi A-L. and Oltvai Z. Nature 411, 41-42 (2001)

Jensen O.N., Curr. Opin. Chem. Biol., 8, 33-41 (2004)

Kaiser M. and Hilgetag CC. Biol. Cybern. 90:311-317 (2004)

Kaiser M, Martin R, Andras P and Young MP. Eur. J. Neurosci.. 25:3185-3192 (2007)

Kitano H. Nat Rev Genet 5, 826-37 (2004)

Kitano H. Molecular Systems Biology 3, 1-7 (2007)

Kola I and Landis J. Nat Rev Drug Discovery 3, 711-716 (2004)

Kola I Clin Pharm and Therapeutics 6100479 (2008)

Krakauer D Complex Systems Science in Biomedicine 2, 183-205 (2006)

Liao J and Laufs U Annual Review of Pharmacology and Toxicology 45, 89-118 (2005)

Masel J and Siegal M Trends in Genetics 25, 395-403 (2009)

Masel J and Trotter M Trends in Genetics 26, 406-414 (2010)

Metz, J.T. and Hajduk, P.J; Curr, Opin. Chem. Biol. 14: 498-504 (2010)

Morphy R J. Med. Chem. 53:1413-1437 (2010)

Morphy, R., Kay C. And Rankovic, Z.; Drug Disc. Today, 9: p 641-651 (2004)

Paolini, G.V., Shapland, R.H., van Hoorn, W.P., Mason, J. and Hopkins.; Nature Biotech.24: 805-815 (2006)

Payne BR et al Proc Natl Acad Sci USA 93, 290-294 (1996)

Piersma T. and Van Gils J. The flexible Phenotype, Oxford Univ Press, pp238 (2011)

Pujol A., Mosca R., Farres J. and Aloy P. Trends in Pharmacological Sciences 31, 115-123 (2010)

Raman K. and Wagner A. Molecular Biosystems 7, 1081-1092 (2011)

Roth, B.L., Sheffler, D.J., Kroeze, W.K., Nature Rev. Drug. Disc. 3: 353-359 (2004)

Scannell JW Blanckley A. The Long View: Pharmaceutical R&D productivity - after six decades of failed cures, isn't it time for a diagnosis? Bernstein Research, Sanford C Bernstein Ltd (2010)

Scannell JW., Anderson T., Redenius J., Bedell Clive L. The Long View: Pharmaceutical R&D productivity follows Moore's Law backwards. Does anyone know why? Bernstein Research, Sanford C Bernstein Ltd (2011)

Scannell JW and Young MP Current Biology 3:191-200 (1993)

Spirin V and Mirny L PNAS 100, 12123-12128 (2003)

Sprague JM Science 153, 1544-47 (1966)

Swinney DC and Anthony J Nature Reviews Drug Discovery 10:507-519 (2011)

Wagner, A. Nat Genet 24, 355-361 (2000)

Wagner, A. and Wright, J. Biosystems 88, 163-172 (2007)

Wagner, A. Proc Roy Soc B 275, 91-100 (2007)

Weitz-Schmidt G et al Nature Medicine 7, 687 - 692 (2001)

Wermuth, C. Drug Disc.Today. 9: 826-827 (2004)

Won J, Chang S, Oh S, Kim TT PNAS 101:11328-11333 (2004)

Young MP Nature 358:152-155 (1992)

Young MP Proceedings of the Royal Society: Biological Sciences 252:13-18 (1993)

Young MP, Hilgetag CC and Scannell JW Philosophical Transactions of the Royal Society: Biological Sciences, 355:147-161 (2000)

Young MP, Andras P and O'Neill MA European Patent Office EP1556819 (2002a)

Young MP, Andras P and O'Neill MA USPTO US7466663 B2 (2002b)

Young MP Drug Discovery World, 3: 9-12 (2008)

Young MP and Whitmore AV. European Biopharmaceutical Review 132:2692 (2010)

Zimmer S and Young MP. Innovation in Pharmaceutical Technology 18: 38-41 (2009)