FEBS Letters 579 (2005) 1872–1877 FEBS 29363

Minireview

Illuminating drug discovery with biological pathways

Gordana Apic[a,*], Tijana Ignjatovic[b], Scott Boyer[b], Robert B. Russell[c]

[a] *Cambridge Cell Networks, William Gates Building, Cambridge CB3 0FD, UK*
[b] *Safety Assessment, AstraZeneca R&D, 43183 Mölndal, Sweden*
[c] *EMBL, Meyerhofstrasse 1, 69112 Heidelberg, Germany*

Accepted 14 February 2005

Available online 21 February 2005

Edited by Giulio Superti-Furga

Abstract Systems biology promises to impact significantly on the drug discovery process. One of its ultimate goals is to provide an understanding of the complete set of molecular mechanisms describing an organism. Although this goal is a long way off, many useful insights can already come from currently available information and technology. One of the biggest challenges in drug discovery today is the high attrition rate: many promising candidates prove ineffective or toxic owing to a poor understanding of the molecular mechanisms of biological systems they target. A "systems" approach can help identify pathways related to a disease and can suggest secondary effects of drugs that might cause these problems and thus ultimately improve the drug discovery pipeline.

Keywords: Systems biology; Drug discovery; Biological pathways

1. Introduction

Since the dawn of medicine, "systems" approaches were used to identify effective treatments. One of the most complex biological systems – a patient with a disease – was effectively treated as a black-box, and exposed to all manner of herbal mixtures and religious rituals. Those treatments showing beneficial effects would be given again to others with similar symptoms, and would eventually become established as drugs. And for thousands of years this process was largely unchanged.

The last fifty years saw the advent of molecular biology and with it modern drug discovery. The observation that drugs arising from herbal remedies typically acted by binding to a single receptor molecule changed the discovery process. To design a drug, one now normally searches for a 'magic bullet' that binds specifically to a rationally chosen target. This reductionist approach has since proved highly effective, and drugs designed in this way have been available for decades. However, it has certain drawbacks. Probably the biggest is that the possibility of a designed molecule binding in places other than the target is often neglected until comparatively late in the discovery process, where many candidate drugs fail. This can, however, have some fortuitous consequences whereby a drug designed for one purpose does something unpredicted, but nevertheless beneficial (e.g., Gleevec [1]).

*Corresponding author. Fax: +49 6221 387 517.
E-mail address: gordana.apic@cambridgecellnetworks.com (G. Apic).

The limits and surprises of the reductionist approach highlight the need to consider the entire system when designing drugs. Fortunately, we are now witnessing breakthroughs in systems biology [2] driven by the latest high-throughput experimental and computational methods. The result is a much better understanding of whole systems, and with it the emergence of possibilities to augment the drug discovery process [3–5]. Here, we discuss a few ways in which systems approaches can already make a difference. Our particular focus is on the use of pathways as a central reference to provide a more holistic view of biological processes relevant to drug discovery.

2. Pathways and pathway resources

The cell can be considered to be a complex network of interacting molecules. Despite many decades of experiments and many thousands of data points, this network is still very far from being complete. Because of its complexity biologists have preferred to consider parts of the cell network, hence the notion of 'pathways', or sets of molecules acting in concert, usually involved in a particular function or process, to the extent that it is convenient to consider them in isolation.

Academic pathway and interaction initiatives focus largely on developing fundamental computational tools to integrate and visualise data from a variety of sources (e.g. [6–8]). Most information comes from high-throughput protein–protein interaction discovery methods (e.g. [9–13]), and as a result tends to be biased towards non-human model organisms: i.e., Yeast, Fly & Worm. For these, there is a growing body of knowledge that is gradually providing a more holistic picture of the organism.

Other initiatives focus on gathering 'textbook' like knowledge using experts and consultation with the literature. There are now over 150 biological pathway databases, with different emphasis and coverage [http://cbio.mskcc.org/prl]. Among the best known are the signaling pathway collections of Biocarta (http://cgap.nci.nih.gov/Pathways/) or STKE (http://stke.sciencmag.org), and metabolic pathways in KEGG (http://www.genome.jp/keg/kegg2.html) [14]. More recently, ambitious projects like Reactome (http://www.reactome.org) have begun to represent pathways as a system of reactions to allow the consideration of variables like time and concentration.

These resources, while extremely useful, have several drawbacks. A major one is incomplete coverage: many parts of the cell network have yet to be studied in sufficient detail,

0014-5793/$30.00
doi:10.1016/j.febslet.2005.02.023

and thus the fraction of proteins currently represented is small. For example, Biocarta contains over 300 pathways in human, but these involve only about 1500 genes from a genome with well over 20000. The missing links can, in principle, be added by considering the results from coarser interactions studies like the two-hybrid or TAP systems, but here one necessarily introduces many new problems, as these data are highly error prone [15,16]. There are also problems related to inconsistencies caused by idiosyncrasies of authors, and of incompatibilities in data formats. Further annotation and experiments to test poorly understood details will gradually remove inconsistencies, and new standards like the Systems Biology Markup Language (SBML) [17] promise for better compatibility. However, it will likely be decades before a complete set of pathways is available.

3. Pathways in drug discovery

Pathway collections have many uses in drug discovery and are already considered by the industry. Here one needs to be pragmatic: questions asked during drug discovery need answers quickly, and methods that provide the most reliable answers given the currently available data are paramount. Pathways can serve as a framework for generating hypothesis for further discoveries, and as such are already useful for many stages of the drug discovery pipeline from identifying novel drug targets to assessing toxicological effects. Pathways put a drug target into context: one can chart those in which a target is seen, and thus make educated guesses about the effects that blocking the target are likely to have. Proteins that are highly connected in interaction networks are much more likely to be lethal when deleted [18] than others, which calls into question the wisdom targeting them to achieve subtle effects; though they may be ideal targets for harsher therapies (e.g., for Cancer) where one wishes to kill over active cells (e.g., the drug glendanamycin, which targets the interaction 'hub' HSP90 [19,20]). One can also look at where proteins similar (i.e., homologous) to the target lie, and make inferences about secondary effects of a drug with imperfect specificity.

Academic resources are not tailored for drug discovery, and this has inspired several commercial providers focusing on pathways relevant to human disease. For instance companies such as Ingenuity Systems, Genego, Ariadne Genomics provide human pathways together with information on various drug action, and concentrate on using pathways for analysis of experimental genomic and proteomic data highly relevant in target discovery and validation. Other providers offer lists of pathway elements for representing pathways in scientific publications (e.g., Protein Launch) with a focus on graphical display. Cambridge Cell Networks (CCNet) concentrates on pathways in pre-clinical drug safety and toxicology. Even though the biological systems are not fully understood, these providers, as well as efforts within large pharmaceutical companies, are beginning to generate useful pathway atlases for studying disease physiology.

For obvious reasons, commercial efforts also place a greater emphasis on integrating chemical data into pathways. This can also be powerful as one can map known drugs or chemical modulators onto pathway collections and exploit chemoinformatics tools, for instance to see where compounds similar to a candidate are known to act within the system.

4. Practical applications: helping to answer questions in drug discovery

Below we pose a number of questions that are typical of the drug discovery process and illustrate how the pathway framework can provide some answers, or suggest further experiments.

4.1. Is a drug target promising?

Target discovery and validation are early, but critical parts of drug discovery. Pathways can be a great aid to testing the validity of a target. At the very simplest level, they can reveal how a target normally behaves – its interconnections in the cell: proteins upstream or downstream of it in signaling cascades or metabolic pathways, etc. One can see the possible knock-on effect of interfering with a target: the downstream effects may be beneficial in one system and detrimental in another. This knowledge can be a great aid when deciding whether a target is worth pursuing: targeting a protein with too many connections might prove disastrous and prompt the search for an alternative. For instance, it may be worth choosing a kinase further downstream to avoid a compound having detrimental effects owing to its action in many pathways.

In many instances, however, drug targets are newly discovered and thus their full biological role is not known. This necessitates constant updates of the connectivity of a target throughout the lifecycle of a drug discovery project. An interesting example of this is cyclooxygenase-2 (Cox-2), first purified in 1976 and cloned in 1988 [21–24]. Its biochemical 'pathway', represented as its role in prostanoid formation from the metabolism of arachidonic acid (Fig. 1) [25], undoubtedly sparked many drug discovery programmes in the late 1980s and early 1990s. This rather simple pathway differs markedly from its term connectivity when mapped using the current literature (Medline; Fig. 1) and represents the challenges of the necessary reductionism of a drug discovery project when the reality of biological complexity considered.

4.2. What side effects might a compound have?

Side effects of candidate drugs are common, and can lead to delays and failures at comparatively late stages in the discovery process. Any effects that can be predicted early are thus of great advantage, and pathways can indeed sometimes help do just this.

Some side effects are caused by a drug binding to a similar (i.e., homologous protein). A good example is the phosphodiesterase (PDE) inhibitor Viagra (Sidenafil). Originally designed to target PDE-5 and promote the relaxation of smooth muscle [26], the compound also binds to the homologous PDE-6 in the eye, which leads to a "blue vision" side effect in patients [27]. Such subtle ocular effects are difficult to detect in animals, thus making efforts to predict them key. CCNet's PathTox tool readily indexes a simple sequence search to our pathway collection and finds these two enzymes and the likely effects of blocking them (Fig.

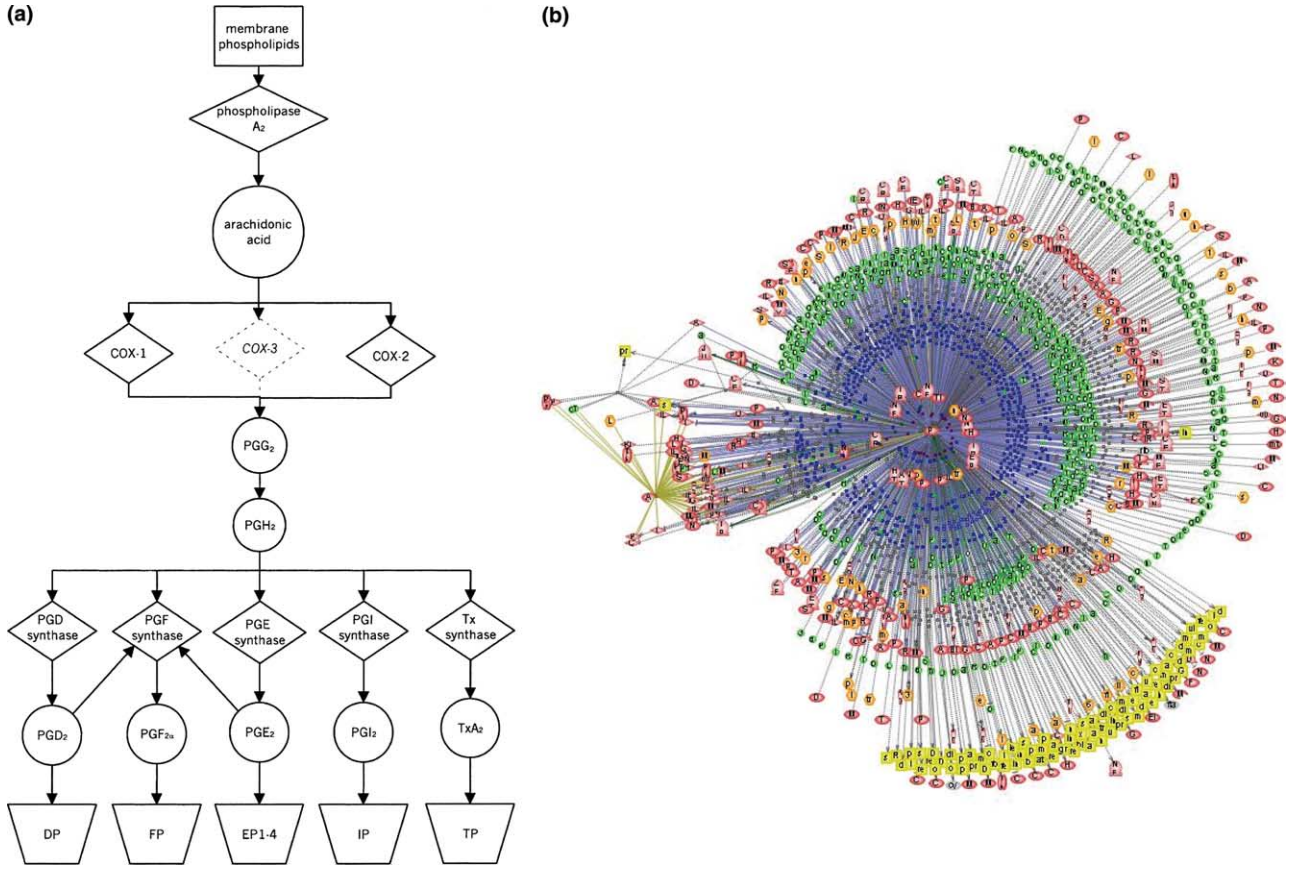

Fig. 1. Pathways involving Cox-2. Biochemical 'Pathway' of Cox-2 [25] (a) in the formation of prostanoids and the term connectivity (b) of Cox-2 as represented by the pathway-mining tool PathwayAssist. Red symbols denote proteins, yellow physiological processes and green represent small molecules.

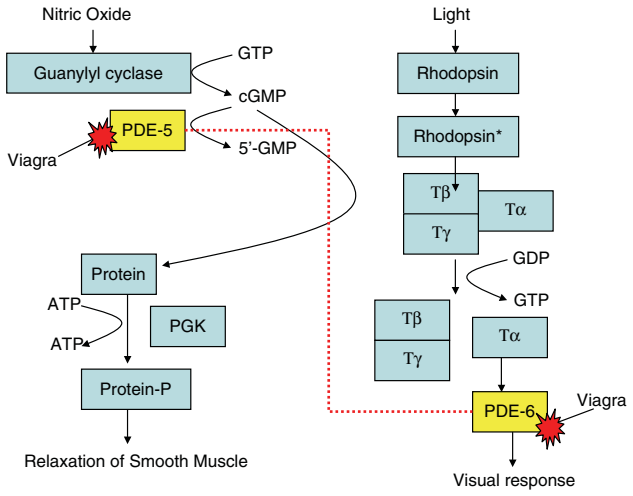

Fig. 2. Example of how drug-specificity problems might be predicted using pathway information. The figure shows simplified parts of the smooth-muscle relaxation pathway (left) and light-sensing pathway in the eye (right). The molecular target for Viagra is PDE-5, a phosphodiesterase in smooth muscle, which is homologous to PDE-6 in the eye (indicated by a broken red line), and to which Viagra (Sidenafil) also binds, leading to a well-documented side effect of blue vision in patients (e.g. [27]).

2). If situations like this are identified early, efforts can be made to design more selective compounds, and thus potentially avoid problems because of this cross-reactivity.

4.3. What is the molecular explanation for an observed toxicity?

Toxicology is currently a great bottleneck. Promising drug candidates very often fail to reach the market owing to unpredictable effects unrelated to the mechanism of drug action (e.g., binding the wrong molecule). Moreover, the molecular basis of toxic effects observed during animal studies are frequently poorly understood, or difficult to transfer to humans. A systems context is clearly needed here [28], and it is here where a pathway framework, when complemented with data from other sources can be of great use.

For instance, if a toxicologist is interested in the possible causes of a physiological observation such as "thyroid hyperplasia", it is desirable to identify candidate molecules and pathways without resorting to days of intensive literature investigation. Here, the pathway scaffold, when cross-referenced to the literature can quickly provide clues. Protein or gene ontologies, annotated features like functional domains or binding sites, or the scientific literature (e.g., PubMed) are of greater use when placed in a pathway context. Trends or similarities can be sought that could give clues as to how pathways overlap or communicate with one another. These come readily when an expert carefully studies the literature and adds information as appropriate, though the process is hard to automate. However, the process can be accelerated with various text searching and mining tools (e.g., Natural Language Processing, NLP) enhanced by biomedical and gene/protein synonyms, or gene (GO) [29] or drug ontologies (e.g., http://www.biowisdom.com). Given query biomedical terms (e.g.,

"bile acids and toxicity") CCNet use the latest NLP technology to retrieve relevant publications, considering not only possible synonyms and ontologies (syntax), but also applying contextual filters (semantics) to remove irrelevant matches. These are then cross-referenced to associated gene products, and then to pathways to give the best possible molecular view of a macromolecular observation (Fig. 3; www.cambridgecellnetworks.com/text).

A key issue in toxicology is to understand whether effects observed in tested animals (e.g., mice, dog) are transferable to humans. Careful discernment of orthologues (i.e., direct equivalents of genes in different species) from paralogues (i.e., genes that may have duplicated since the species diverged) is thus critical [30]. For instance, it is conceivable that one might be able to dismiss a toxic effect observed, say, in mouse if one can argue that corresponding pathways are different in humans, thus rescuing a compound that might otherwise be discarded.

4.4. How can toxic effects be tracked indirectly?

Many toxic effects are related to a candidate molecule binding to proteins other than the target which are involved in critical cell processes. Even when the mechanism is well understood, the biochemistry or the location of the molecules involved can make direct testing for the phenomenon difficult. A systems view, in particular 'pathway walking' can often uncover an alternative, indirect test.

We have applied this approach in order to suggest alternative routes for testing whether candidate molecules interfere with bile acid transport (a common toxic effect). Bile acid homeostasis in the liver is tightly controlled by a system involving several nuclear hormone receptors (FXR, LXR, PXR, etc.) [31,32]. Bile acids activate FXR, which in turn regulates the expression of enzymes involved their synthesis (Cyp7A1 and Cyp8B1) [33,34], through a negative feedback loop. FXR also regulates the expression of several bile-acid transport proteins including BSEP/SPGP [35]. A full picture of this process can be obtained only through a combination of metabolic and signaling pathways and the extraction of genes regulated by FXR using text-mining (Fig. 4).

Assaying the biochemical activity of transporters directly is extremely problematic. However, pathway-walking along the gene regulatory network surrounding FXR suggests that transporter activity would be tightly correlated with the expression of other FXR regulated genes, and thus in principle one could monitor transport function indirectly via expression analysis.

5. Combination drugs: towards the design of 'magic shrapnel'

Most of the discussion above, like much of modern drug discovery, assumes that one is dealing with a single compound with (in the best scenario) a single target. However, this rather neglects both ancient and modern remedies that consist of two or more compounds acting in concert.

Treatments containing more than one bioactive compound are as old as medicine itself. Herbal remedies have long been known to contain a great number of different substances, and it is often difficult to discern what, if anything, each of them is doing. There are some indications that herbal remedies elicit their beneficial effects by tinkering with different receptors in pathways in a gentle way. Instead of a single, concentrated 'magic bullet', they contain several compounds with lower concentrations more like 'magic shrapnel'. It is even possible for one compound to partly counteract another. For instance, the nutracuetical ginseng has both wound healing and antitumor effects through opposing activities on vascular system [36]. DNA microarrays and chemical fingerprinting have recently revealed that ginseng indeed contains two active substances with opposite effects, Rg1 that leads to angionenesis

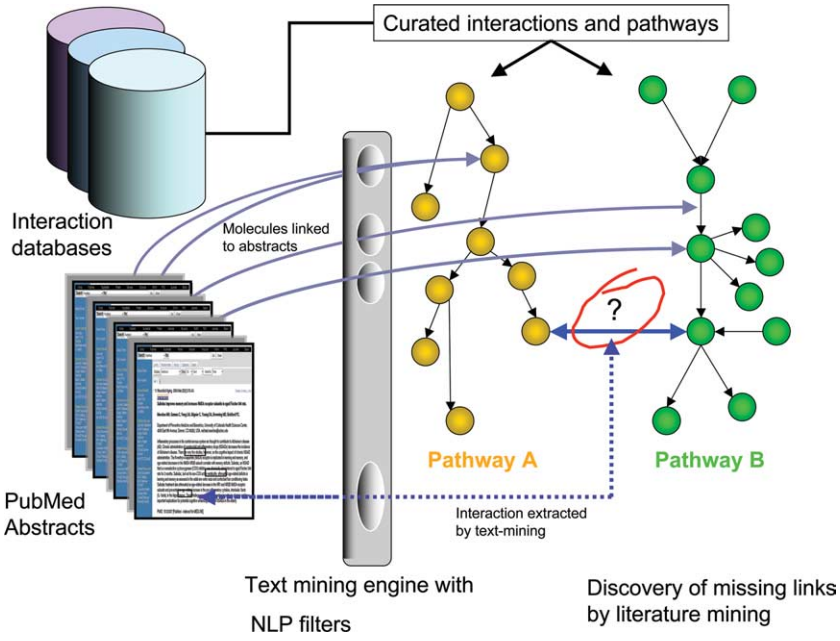

Fig. 3. Integration of pathways, interactions and the literature to find missing links. The figures shows how a text query can be used (via NLP) to extract meaningful abstracts and their associated molecules and interactions, and how these can be mapped onto pathways (gray lines) to suggest missing links between pathways (blue line), and to provide explanations for the molecular mechanisms related to the query.

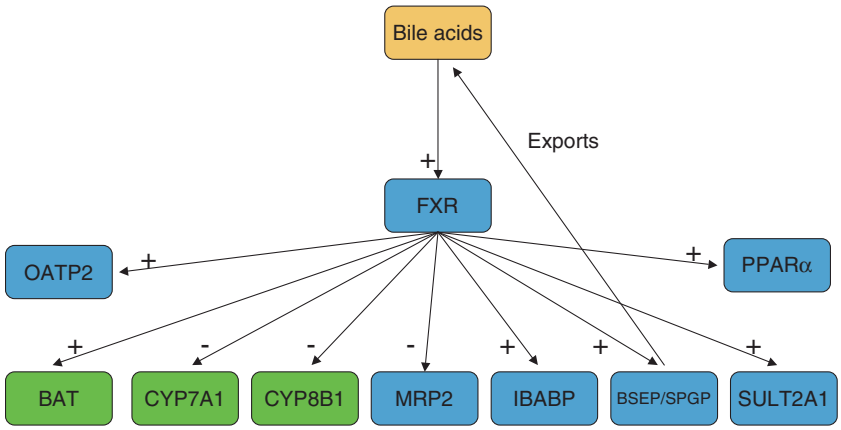

Fig. 4. Regulation of genes by FXR in the presence of bile acid. The figure shows a part of the negative feedback loop in control of bile acid homeostasis in liver. Enzymes involved in bile acid synthesis are shown in green and the other proteins in blue (farnesoid X receptor, FXR), organic anion transport protein 2 (OATP2), DHEA–sulfate transferase (SULT2A1), peroxisome proliferator-activated receptor-alpha (PPARα), ileal bile acid biding protein (IBABP), multidrug resistance protein 2 (MRP2), Sister of P-glycoprotein SPGP/bile salt export pump (BSEP/SPGP). Plus/minus symbols (+/−) denote increase/decrease in expression or activity through direct or indirect mechanisms. The elements of this pathway were extracted using CCNet's software.

via the Akt Pathway and Rb1 that prevents it acting in the early angiogenesis pathway [36].

This natural precedent poses the obvious question for new drugs: can we rationally design mixtures such as those discovered in herbal medicine? Probably a deeper understanding of pathways is ultimately needed, but certainly great hope comes from the revolutionary approach of CombinatoRx. Rather than attempting rational design based on a single drug target, they performed a high-throughput screen of 120 000 pair-wise combinations of 600 established drugs in various phenotypic assays (from cellular assays to whole organisms such as zebra fish) [37]. The hope was that side effects typical of new drugs could be avoided through the use of combinations of those that have already passed successfully through toxicology, safety and clinical trials. The results were astonishing: some combinations had unpredictable beneficial effects that were not related to those known for the individual substances. For instance, the combination of the anti-psychotic chlorpromazine and the anti-protozoal pentamidine, neither of which show any anti-tumor activity, prevented the tumor growth in mice [37]. No obvious mechanism was clear from what was known about the two compounds: a new effect resulted from a delicate interplay between the pathways involved.

Screens like these provide unprecedented insight into connectivites of biological pathways. The challenge for experimental and computational systems biologists is to make sense of such complex mixtures beyond what is already known about their individual components.

6. Concluding remarks

Scientists in pharmaceutical companies are continually forced to make tough decisions related to which of several compounds should be taken further when millions have already been spent on each of them. The decisions must often be made without a detailed understanding of the mechanism of action. Better understanding of human disease biology will clearly lead to faster and more accurate decisions. Biological pathways, though still error-prone and far from complete in coverage, can now provide a more efficient way of browsing through pharmaceutically relevant information, and offer a quick overview of underlying molecular processes. These and other efforts in systems biology are already progressing towards a systematic understanding of the molecular mechanisms of disease biology and drug action.

References

[1] Pardanani, A. and Tefferi, A. (2004) Imatinib targets other than bcr/abl and their clinical relevance in myeloid disorders. Blood 104, 1931–1939.
[2] Ideker, T., Galitski, T. and Hood, L. (2001) A new approach to decoding life: systems biology. Annu. Rev. Genom. Hum. Genet. 2, 343–372.
[3] Butcher, E.C., Berg, E.L. and Kunkel, E.J. (2004) Systems biology in drug discovery. Nat. Biotechnol. 22, 1253–1259.
[4] Werner, E. (2002) Systems biology: the new darling of drug discovery?. Drug Discov. Today 7, 947–949.
[5] Davidov, E., Holland, J., Marple, E. and Naylor, S. (2003) Advancing drug discovery through systems biology. Drug Discov. Today 8, 175–183.
[6] Bader, G.D., Betel, D. and Hogue, C.W. (2003) BIND: the biomolecular interaction network database. Nucleic Acids Res. 31, 248–250.
[7] Xenarios, I., Salwinski, L., Duan, X.J., Higney, P., Kim, S.M. and Eisenberg, D. (2002) DIP, the database of interacting proteins: a research tool for studying cellular networks of protein interactions. Nucleic Acids Res. 30, 303–305.
[8] Zanzoni, A., Montecchi-Palazzi, L., Quondam, M., Ausiello, G., Helmer-Citterich, M. and Cesareni, G. (2002) MINT: a molecular interaction database. FEBS Lett. 513, 135–140.
[9] Uetz, P., et al. (2000) A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. Nature 403, 623–627.
[10] Rain, J.C., et al. (2001) The protein–protein interaction map of *Helicobacter pylori*. Nature 409, 211–215.
[11] Gavin, A.C., et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415, 141–147.
[12] Giot, L., et al. (2003) A protein interaction map of *Drosophila melanogaster*. Science 302, 1727–1736.

[13] Li, S., et al. (2004) A map of the interactome network of the metazoan *C. elegans*. Science 303, 540–543.
[14] Kanehisa, M. (2002) The KEGG database. Novartis Found. Symp. 247, 91–101, (discussion 101–3, 119–28, 244–52).
[15] von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S. and Bork, P. (2002) Comparative assessment of large-scale data sets of protein–protein interactions. Nature 417, 399–403.
[16] Aloy, P. and Russell, R.B. (2002) Potential artefacts in protein-interaction networks. FEBS Lett. 530, 253–254.
[17] Hucka, M., et al. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19, 524–531.
[18] Jeong, H., Mason, S.P., Barabasi, A.L. and Oltvai, Z.N. (2001) Lethality and centrality in protein networks. Nature 411, 41–42.
[19] Scheibel, T. and Buchner, J. (1998) The Hsp90 complex – a super-chaperone machine as a novel drug target. Biochem. Pharmacol. 56, 675–682.
[20] Stanyon, C.A. and Finley Jr., R.L. (2000) Progress and potential of *Drosophila* protein interaction maps. Pharmacogenomics 1, 417–431.
[21] Yokoyama, C., Takai, T. and Tanabe, T. (1988) Primary structure of sheep prostaglandin endoperoxide synthase deduced from cDNA sequence. FEBS Lett. 231, 347–351.
[22] Merlie, J.P., Fagan, D., Mudd, J. and Needleman, P. (1988) Isolation and characterization of the complementary DNA for sheep seminal vesicle prostaglandin endoperoxide synthase (cyclooxygenase). J. Biol. Chem. 263, 3550–3553.
[23] DeWitt, D.L. and Smith, W.L. (1988) Primary structure of prostaglandin G/H synthase from sheep vesicular gland determined from the complementary DNA sequence. Proc. Natl. Acad. Sci. USA 85, 1412–1416.
[24] Vane, J.R., Bakhle, Y.S. and Botting, R.M. (1998) Cyclooxygenases 1 and 2. Annu. Rev. Pharmacol. Toxicol. 38, 97–120.
[25] Warner, T.D. and Mitchell, J.A. (2004) Cyclooxygenases: new forms, new inhibitors, and lessons from the clinic. FASEB J. 18, 790–804.
[26] Kuthe, A., Montorsi, F., Andersson, K.E. and Stief, C.G. (2002) Phosphodiesterase inhibitors for the treatment of erectile dysfunction. Curr. Opin. Invest. Drugs 3, 1489–1495.
[27] McCulley, T.J., Lam, B.L., Marmor, M.F., Hoffman, K.B., Luu, J.K. and Feuer, W.J. (2000) Acute effects of sildenafil (viagra) on blue-on-yellow and white-on-white Humphrey perimetry. J. Neuroophthalmol. 20, 227–228.
[28] Bugrim, A., Nikolskaya, T. and Nikolsky, Y. (2004) Early prediction of drug metabolism and toxicity: systems biology approach and modeling. Drug Discov. Today 9, 127–135.
[29] Ashburner, M., et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25, 25–29.
[30] Chervitz, S.A., et al. (1998) Comparison of the complete protein sets of worm and yeast: orthology and divergence. Science 282, 2022–2028.
[31] Chiang, J.Y. (2004) Regulation of bile acid synthesis: pathways, nuclear receptors, and mechanisms. J. Hepatol. 40, 539–551.
[32] Sinal, C.J., Tohkin, M., Miyata, M., Ward, J.M., Lambert, G. and Gonzalez, F.J. (2000) Targeted disruption of the nuclear receptor FXR/BAR impairs bile acid and lipid homeostasis. Cell 102, 731–744.
[33] Chen, W., Owsley, E., Yang, Y., Stroup, D. and Chiang, J.Y. (2001) Nuclear receptor-mediated repression of human cholesterol 7alpha-hydroxylase gene transcription by bile acids. J. Lipid Res. 42, 1402–1412.
[34] Chiang, J.Y., Kimmel, R. and Stroup, D. (2001) Regulation of cholesterol 7alpha-hydroxylase gene (CYP7A1) transcription by the liver orphan receptor (LXRalpha). Gene 262, 257–265.
[35] Schuetz, E.G., et al. (2001) Disrupted bile acid homeostasis reveals an unexpected interaction among nuclear hormone receptors, transporters, and cytochrome P450. J. Biol. Chem. 276, 39411–39418.
[36] Sengupta, S., et al. (2004) Modulating angiogenesis: the yin and the yang in ginseng. Circulation 110, 1219–1225.
[37] Borisy, A.A., et al. (2003) Systematic discovery of multicomponent therapeutics. Proc. Natl. Acad. Sci. USA 100, 7977–7982.