FEBS Letters 582 (2008) 1283–1290

Minireview

Large-scale prediction of drug–target relationships

Michael Kuhn[a,1], Mónica Campillos[a,1], Paula González[a,1], Lars Juhl Jensen[a,b,1], Peer Bork[a,c,*,1]

[a] European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany
[b] The Novo Nordisk Foundation Centre for Protein Research, University of Copenhagen, Blegdamsvej 3, 2200 Copenhagen, Denmark
[c] Max-Delbrück-Centre for Molecular Medicine, Robert-Rössle-Strasse 10, 13092 Berlin, Germany

Received 21 December 2007; revised 8 February 2008; accepted 11 February 2008

Available online 20 February 2008

Edited by Patrick Aloy and Robert B. Russell

Abstract **The rapidly increasing amount of publicly available knowledge in biology and chemistry enables scientists to revisit many open problems by the systematic integration and analysis of heterogeneous novel data. The integration of relevant data does not only allow analyses at the network level, but also provides a more global view on drug–target relations. Here we review recent attempts to apply large-scale computational analyses to predict novel interactions of drugs and targets from molecular and cellular features. In this context, we quantify the family-dependent probability of two proteins to bind the same ligand as function of their sequence similarity. We finally discuss how phenotypic data could help to expand our understanding of the complex mechanisms of drug action.**

Keywords: Drug–target interaction; Small-molecule network; Drug discovery; Target prediction

1. Introduction

The increasing amount of publicly available chemical data creates opportunities for the analysis and integration of resources of molecular information at the interface between biology and chemistry. While large-scale data sets have long been publicly available in molecular biology, this spirit of openness began only recently to spread in chemistry. Funding bodies such as the National Institutes of Health (NIH) are fostering the creation of public databases, for example, PubChem [1] as part of the NIH's Molecular Libraries Roadmap Initiative. In addition, more research areas are being considered pre-competitive by the pharmaceutical industry. Consequently, we are witnessing an increasing number of public databases that store information about compounds along with properties and context.

The combined knowledge on individual drugs and targets can be advantageously integrated with new high-throughput data sets and concepts for systems-wide analysis of their relations, thus opening a new road to predict drug–target relationships and the effects of drugs on human biology. Until exhaustive screens have been performed that study the effect of all human drugs on all human proteins under various conditions [2,3], computational and systems biology approaches will be invaluable in extending our knowledge on drug–target relations systematically.

Here we (i) review publicly available resources of known drug–target relations with the aim to define a gold standard of 'positives' for benchmarking predictive approaches, (ii) illustrate how a global network view provides essential context for individual drug–drug and drug–target relations, (iii) discuss molecular features of drugs and target proteins that can be utilized for the prediction of drug–target relations, and finally, (iv) describe how phenotypic information could help to expand our understanding of the molecular and cellular effects of drugs. Although we focus here on relations between drugs and targets, many of the presented approaches and resources are applicable to chemical–protein relations in general. Likewise, chemical–protein relations implicitly include those of drugs and their targets. We deliberately do not address the impact of individual genetic makeup [4,5] and environmental factors [6,7] on both mechanism of action and toxicity of drugs as the amount of available data is still very limited.

The exploitation and integration of heterogeneous data from existing resources will enable the prediction of many hitherto unknown targets for existing drugs eventually resulting in new leads for treating human diseases. The inclusion of the context of individual drug–target relations, e.g. in the form of a network, will also aid in anticipating indirect consequences of drug treatment such as side effects and undesirable drug interactions.

2. Resources and approaches for large-scale prediction and analysis of protein–chemical relations

2.1. Capturing the existing knowledge

To form a basis for the prediction of novel drug–target relationships it is necessary to collect as much information as possible on small molecules, proteins and their interactions. Historically, chemists and biologists have taken very different approaches to storing and sharing data.

Information on the sequence, structure and function of proteins is collected in public databases such as UniProt [8] and PDB [9]. By contrast, chemical databases have traditionally been commercial and thus not freely accessible. Public databases on proteins emerged in the 1970s, fostered by the availability of digital storage and by requirements from publishers to deposit data in public resources. The history of chemical databases can be traced into the 19th century; for example,

*Corresponding author. Address: European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany.
E-mail address: bork@embl-heidelberg.de (P. Bork).

[1]All authors contributed equally to the manuscript.

0014-5793/$34.00
doi:10.1016/j.febslet.2008.02.024

the Beilstein Handbook of Organic Chemistry has been published since 1881. The distribution of data in the form of books and the economic success of the chemical industry have lead to a tradition of commercial databases on chemical structures and their properties such as the Chemical Abstracts Registry. Only in the past decade several public alternatives have been created, including repositories like PubChem [1], ChEBI [10] and ChemDB [11] that contain information on chemicals and their physicochemical properties. Other databases such as ZINC [12] have been designed as resources for virtual screening applications. These emerging public databases allow access to useful parts lists of proteins and chemicals.

For understanding higher-order processes these parts list have to be connected by determining how the parts interact within biological systems. For proteins, several public repositories for experimentally determined interactions have been established (e.g. BioGRID [13], IntAct [14] and MINT [15]) using a common standard, PSI-MI [16]. Notably, publishers enforce that new experimental evidence is openly accessible to the research community, for example data from high-throughput screens for physical [17,18] and genetic interactions [19,20]. This provides a foundation for the construction of tools that integrate such interactions with other data types (e.g. the STRING database [21] and other resources reviewed in [22]).

The corresponding databases for the relationships of chemicals have not yet reached a comparable state. Although large-scale screens of chemicals in cell-based assays have been performed and are available from repositories such as ChemBank [23] and PubChem BioAssay, deposition of data from chemical screens in standardized repositories is not being enforced. The difficulties involved in obtaining and combining the data has hampered the development of methods for predicting relationships between drugs; to our knowledge currently only one public tool exists that combines data from chemical screens and other sources to infer relationships for chemicals, namely STITCH [21] (Fig. 1).

Databases that centre on drug–target relations are also emerging in the public sector: the Therapeutic Target Database TTD [24], DrugBank [25], SuperTarget [26] and Matador [26] all collect direct drug–target interactions. In addition, Matador [26] includes indirect drug–target interactions that capture more distant effects of drugs on the human protein network. Resources like the PDSP K_i database [27] and BindingDB [28] provide in vitro binding affinities that add knowledge about potential lead molecules; for example, Roth and collaborators discovered that Salvinorin A, the main active ingredient of the hallucinogenic plant *Salvia divinorum*, is a potent kappa opioid agonist by screening it on a collection of receptors [29]. The accumulated content of these databases (summarized in Table 1) constitutes a gold standard. Such a standard is crucial for the development of prediction methods, for example, in the context of proper benchmarking protocols.

All the databases described above contain experimental data related to individual proteins, chemicals or binary interactions. To obtain a global picture of their interplay, the data therein can be integrated with a variety of existing molecular, cellular and organismal data such as microarray experiments (e.g. GEO [30] or ArrayExpress [31]) and pathways (e.g. Reactome [32], KEGG [33] or MetaCyc [34]). By bringing together these heterogeneous data types, it is possible to construct a network that captures many aspects of how drugs and other small-molecules function in a cellular context; for an example see Fig. 1 created using the STITCH database and its visualization capabilities [21].

2.2. Context and its visualization

Systems biology approaches are increasingly being applied to investigate the relationships between proteins, utilizing the biological context of a protein to gather more information about its function [22]. Similarly, the context of a drug needs also to be considered as drugs usually do not only affect the action of isolated targets, but influence entire pathways. Thus the introduction of systems biology concepts into drug discovery is being foreseen [35,36]. While there are many specialized tools to visualize the context of proteins (e.g. [37–40]), chemists mostly have to resort to general purpose tools such as Cytoscape [41,42] to view networks involving chemical compounds, although first visualization tools are emerging [21].

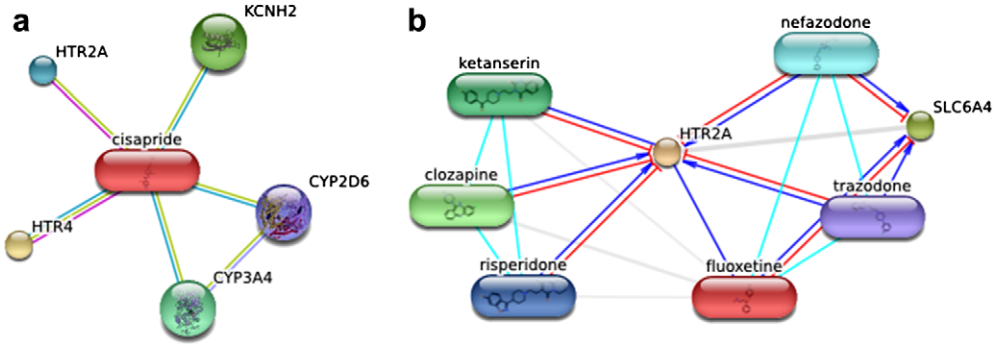

Fig. 1. Network context of drugs and targets. Proteins are shown as spheres (with representative PDB structures, if available) and chemicals as capsules. Connecting lines (edges) depict known or predicted associations. Edge representation depends on query and visualization mode of the STITCH resource [21] from which both examples are taken. (a) Drug–target relationships of cisapride. The serotonin receptor (HTR4 and HTR2A) agonist cisapride also binds to the cardiac ion channel hERG (KNCH2), which leads to arrhythmias as a side effect. In addition, the network shows the interaction of cisapride with metabolizing Cytochrome P450 enzymes (CYP3A4 and CYP2D6). The interactions are derived from various sources, as depicted by the colored lines: experiments (magenta), databases (cyan), text mining (yellow) and homology (lavender). (b) Antagonists of serotonin receptors and serotonin transporter inhibitors. Compounds with similar MeSH (Medical Subject Headings) pharmacological action are connected by cyan lines and form two distinct groups. The *serotonin receptor antagonists* ketanserin, clozapine and risperidone bind (blue line) and inhibit (red line) the serotonin receptor HTR2A. By contrast, the *second-generation antidepressive agents* fluoxetine, nefazodone and trazodone are known to be more promiscuous and inhibit both the serotonin transporter SLC6A4 and the serotonin receptor HTR2A.

Table 1
Databases freely available for academic research that contain information on drug–target interactions

Database	Number of chemicals	Content
Ligand–target databases		
DrugBank http://redpoll.pharmacy.ualberta.ca/drugbank/	~1000 FDA-approved drugs, and ~3000 experimental drugs	6000 drug–targets relationships; chemical, pharmacological and pharmaceutical data
Matador http://matador.embl.de/	~770 drugs	~7000 direct and ~5000 indirect drug–target relationships; links to literature sources for interactions
SuperTarget http://insilico.charite.de/supertarget/	~1500 drugs	7300 drug–target relations
TTD http://bidd.nus.edu.sg/group/cjttd/TTD_ns.asp	~2100 drugs	drug–target relationships with 1535 targets
PDSP K_i http://pdsp.med.unc.edu/pdsp.php	~6800 chemicals	~46,000 K_i values
BindingDB http://www.bindingdb.org/	~18 000 chemicals	~30 000 records with K_i, IC50, or thermodynamic data
Cellular assays		
PubChem BioAssay http://pubchem.ncbi.nlm.nih.gov/	~560 000 chemicals	~600 single compound and high-throughput screening assays
ChemBank http://chembank.broad.harvard.edu/	~1.2 million chemicals	2500 high-throughput biological assays from 188 screening projects

Note: K_i: inhibition constant; IC50: concentration of an inhibitor that is required for 50% inhibition of its target.
Database content was recorded on November 1, 2007.

Networks have been used recently to analyse topological and global properties of chemical–protein interactions such as polypharmacology (a term usually used to describe multiple actions for the same drug [43], for examples, see Fig. 1) and drug target–disease relationships [44,45]. Paolini and co-workers [46] provided an overview of the existing polypharmacology relations by integrating data from several proprietary and public chemical screening sources. The authors presented a protein network in which two proteins are connected if chemicals are known to bind both of them with similar affinity. In this network a highly family-dependent degree of promiscuity of targets was observed, both within the same family and across families.

A similar level of target promiscuity was observed by Yıldırım and co-workers [45] in a network of known drug–target relationships obtained from DrugBank [25]. Within a interaction network of human proteins derived from yeast two-hybrid screens [47,48], the authors also explored the distribution of distances between drug targets and disease genes described in OMIM [49]. Although some drugs were found to target disease genes or their direct network neighbours, the distance distribution otherwise matched that of the random control. This suggests that most drugs in fact alleviate the symptoms (being palliative drugs) rather than target directly the actual cause of the disease [45].

Although such networks attempt to offer a global view on the relations of proteins and chemicals, our knowledge of drug–target relations is far from complete and needs to be expanded in order to increase our understanding of the actions of drugs. One promising avenue in this regard is the accurate prediction of drug–target relations followed by directed experimental validations.

3. Concepts for large-scale drug–target predictions

3.1. Predicting relations based on molecular features of chemicals and proteins

Exploiting similarities between chemical structures is a common way to infer the activity of compounds. The most prevalent approach for comparing compounds is to convert the two-dimensional representation of each compound into a fingerprint either by using a defined list of substructures or by encoding (hashing) all the encountered substructures up to a certain size. This results in fixed-length bit vectors for which the Tanimoto (or Jacquard) similarity measure is computed by dividing the size of intersection of the set bits by the size of the union [50]. Alternatively, chemical similarity can be determined by aligning three-dimensional models of the compounds [51–53]. To illustrate these similarity measures, we show two- and three-dimensional structure comparisons of the monoamine oxidase inhibitor pargyline with five other compounds (Fig. 2).

Initial optimistic results [54] on the relationship between chemical similarity and activity were put into perspective by the analysis of more unbiased chemical libraries. For these, there is only a 30% chance of binding the same compound at the similarity level previously thought to warrant >80% chance [55]. For example, only one of the compounds in Fig. 2 with high similarity to pargyline also inhibits monoamine oxidase. To overcome the limited predictive power of pairwise chemical structure comparison, Keiser and co-workers developed a statistical model to detect remote, yet significant similarities between groups of drugs and used it to predict novel drug–target relations [56]. Other groups used Bayesian classifiers

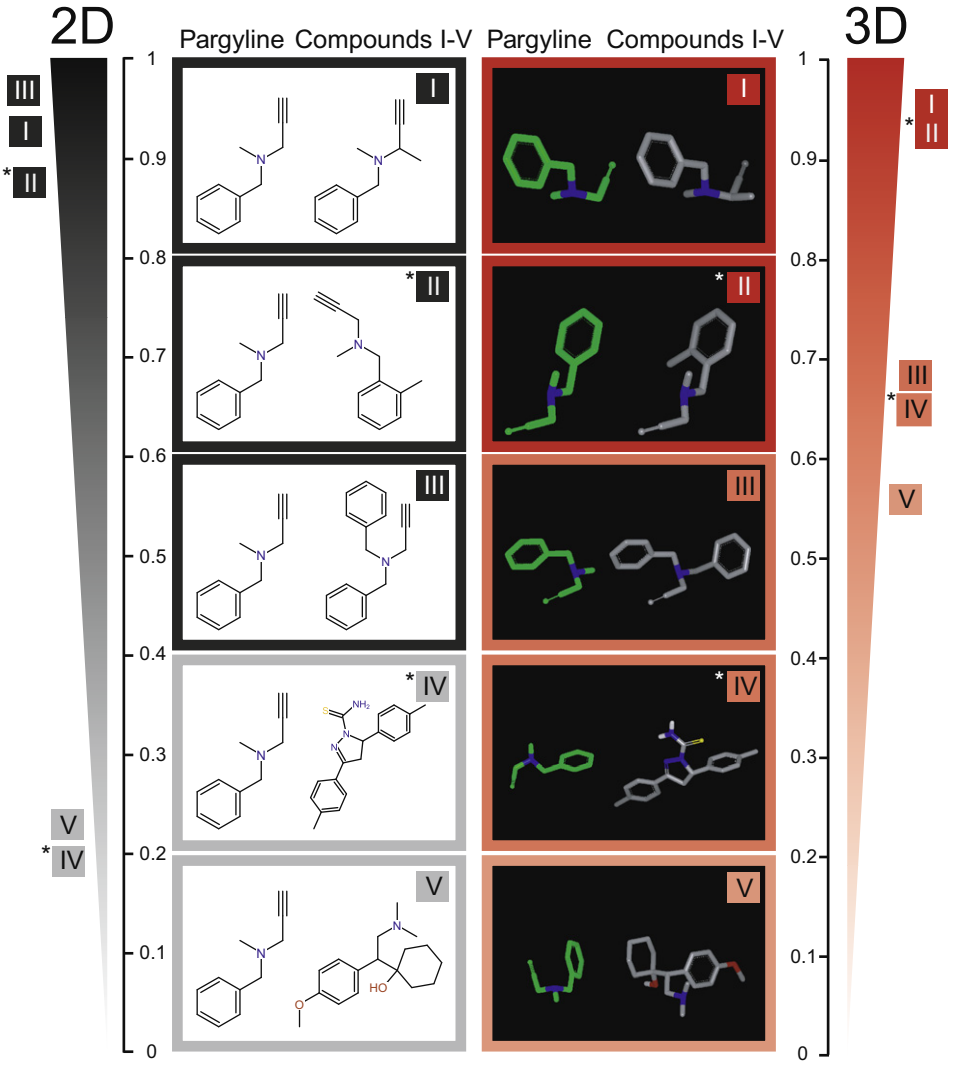

Fig. 2. Comparison of chemical similarity measures. The structure of monoamine oxidase inhibitor pargyline is compared against three pargyline derivatives (compounds I–III, [55]), 1-thiocarbamoyl-3,5-di-(4-methylphenyl)-4,5-dihydropyrazole (compound IV) and Venlafaxine (compound V) [27]. The three-dimensional chemical structures in each panel show the conformation of maximum spatial overlap between the two compounds. Compounds that show activity in a monoamine oxidase inhibition assay [55] are marked with an asterisk. 2D fingerprints and Tanimoto scores were calculated with the Chemistry Development Kit [87]. 3D Tanimoto scores were computed by creating conformers with OMEGA [88] and subsequent shape comparison with ROCS [88].

to correlate the presence or absence of chemical substructures with protein binding properties and reported high success rates for known interactions [46,57,58]. More specialized chemical similarity methods have also been developed that take, for example, the similarity of target proteins into account [59].

Homology relations between proteins can be exploited to predict binding of drugs to proteins that are related to known drug targets [2]. A study on crystal structures of alpha-helical proteins in the PDB showed that the chemical similarity between ligands is higher for proteins with similar sequences [60]. Here, we generalise this to all proteins for which ligand binding constants are available from the PDSP K_i database [27]. Using $K_i = 10\ \mu M$ as the threshold for what is considered "binding", we quantify the probability that two proteins bind the same ligand as a function of their sequence similarity separately for four classes of target proteins (Fig. 3).

Considerable predictive power is observed for G-protein coupled receptors (GPCRs), the largest class of proteins in the database. The probability of binding the same ligand is close to zero for proteins without detectable similarity, but increases to over 60% at a normalized bitscore of about 0.2 (on average corresponding to about 30% sequence identity, see Fig. 3). From a target-prediction perspective, it is thus likely that two drugs cross-react with their GPCR targets only if the sequences of latter are recognizably similar to each other. A similar, albeit less prominent trend is observed for nuclear receptors and for non-kinase enzymes. By contrast, the probability of two protein kinases (including receptor tyrosine kinases) to bind the same ligand remains almost constant (around 10–30%) throughout the range of their respective protein similarities. While evolutionary distant GPCRs and enzymes (other than protein kinases) have a very low probability of sharing the same ligand, this is not the case for homologous kinases with low sequence similarity. These findings agree with previous studies [61,62], which found that kinase inhibitors show little specificity towards similar proteins

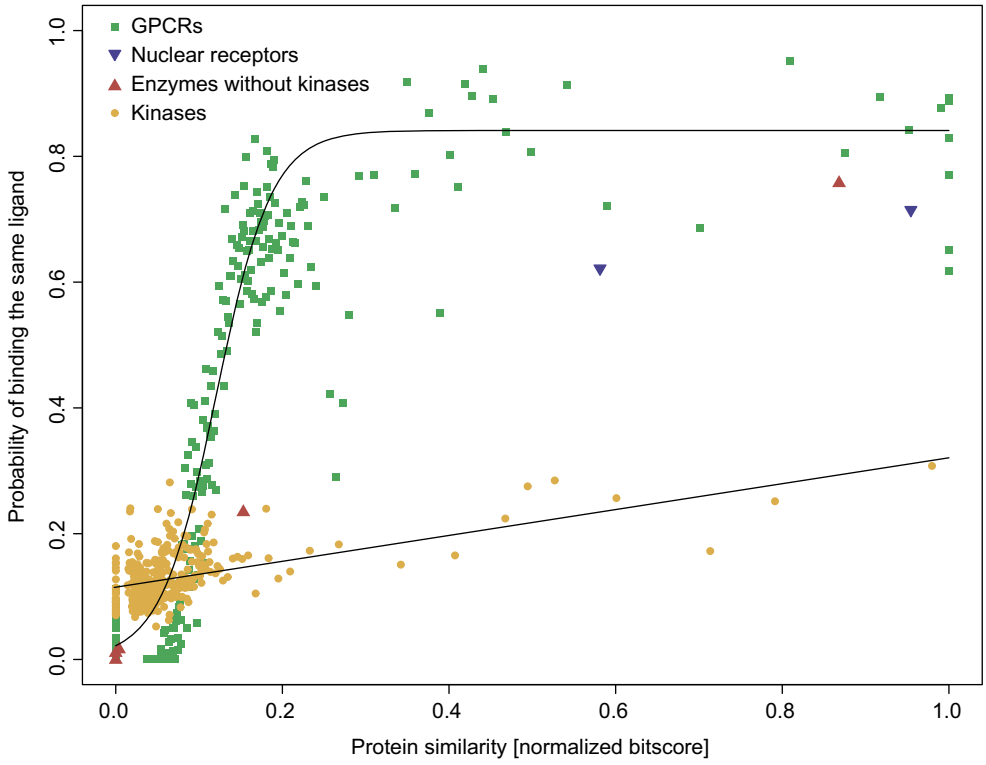

Fig. 3. Protein similarity vs. binding. For different classes of proteins as derived from Gene Ontology categories, the probability of two proteins to bind the same ligand is shown at different levels of protein similarity. Proteins were aligned using the Paralign implementation of the Smith–Waterman algorithm [89] and bitscores were normalized by dividing the bitscore of the alignment by the maximum bitscore achieved by aligning each of the proteins against itself. Only a very weak correlation is observed for kinases (including receptor tyrosine kinases). By contrast, other enzymes and receptors, in particular G-protein coupled receptors (GPCRs), have a high probability of binding the same ligands if the normalized bitscore is 0.2 or higher (corresponding on average to >30% sequence identity).

(see Fig. 3 of Fabian et al. [62]). The experimental results by Fabian and co-workers [62] currently comprise most of the data on kinases in the PDSP K_i database.

While sequence similarity measures can already directly be used to predict ligand sharing for two proteins, spatial molecular features should also be considered. In the case of the kinase protein family, non-polar residues surrounding the ATP binding site and their dehydration propensity hot spots seem to determine binding promiscuity and specificity [63]. This illustrates the importance of taking into account the three-dimensional structure of drugs and their targets.

Protein and chemical structure matches are yet another category of molecular features that can be utilized for drug–target predictions. These three-dimensional fits are usually exploited for lead discovery and optimization by using a variety of docking strategies for computational virtual screening. In addition to the structure of the biomolecular target, all docking algorithms require two components: a scoring function and a search method to find its optimum [64]. Docking of known or constructed compounds has been used to discover novel ligands for well over 30 targets (see [65] for examples) and it has also revealed novel activities of marketed drugs. For example, a recent screen revealed that phenothiazine antipsychotics are weak antagonists of the human androgen receptor. Further optimization of this new lead improved their antagonist effect on the androgen receptor and reduced the effects of their primary target [66].

3.2. Exploitation of phenotypic effects of drug treatment for the prediction of drug–target relations

Phenotypic information from diseases has been valuable to predict novel associations between genes and diseases (e.g. [44,67]). In the context of small molecules, information from phenotypic assays has been used extensively to find lead therapeutic compounds [68–70] and more recently has been exploited computationally by phenotypic profiling methods to predict novel chemical–chemical and chemical–gene associations [71,72].

In phenotypic profiling methods, each compound is screened against a battery of phenotypic assays. The resulting activity profile can subsequently be compared with those of other compounds to infer novel chemical–chemical relationships. Three types of profiles are commonly used: gene-expression, cytotoxicity and chemical–genetic profiles. Gene-expression profiling methods compare the changes in gene expression upon treatment with chemicals to predict which chemicals may have a common mechanism of action [73,71]. Using such an approach, novel relationships between genes, chemicals, pathways and diseases can also be found, for example by comparing gene-expression profiles for chemicals with those for gene mutations [71] or disease states [74].

Cytotoxicity-profiling methods record the growth inhibition of cell lines caused by treatment with compounds. Cytotoxicity profiles across 60 human tumour cell lines (NCI60) have been analyzed extensively by the US National Cancer Institute

(NCI), for example, finding correlations between known mechanism of action and activity profiles [72] or gene expression [75]. Chemical–genetic profiling methods exploit the enhanced drug sensitivity of diploid yeast cells in which the copy number of a target gene is reduced. From the growth inhibition caused by each compound in a collection of yeast haploid deletion mutants [76], a profile is derived. Chemicals with similar profiles have been shown to have similar target activity [77,78].

Correlating similar biological activities between different compounds makes it possible to discover relations between chemicals with the same mechanism of action and thus to make inferences about the targeted proteins or pathways. For instance, the similarity of chemical–genetic profiles of amiodarone (an antianginal and antiarrhythmic) and tamoxifen (a breast cancer therapeutic) observed in a yeast haploid deletion screening suggests that tamoxifen disrupts calcium homeostasis as amiodarone does [79].

This logic can be extended to other phenotypic measures. For example, drugs with similar biological activity show similar side effects [80]. Therefore, by comparing side effects profiles of drugs, it is likely that novel associations between drugs and protein targets can be found. In addition to drug side effects, other types of phenotypic data like drug interactions, when combined with phenotypic information from knockout animal models, could possibly be exploited in a large-scale manner to find novel associations between drugs, their targets and possibly the disease pathway in which the drugs are involved [81].

4. Conclusions

The complex molecular, cellular and organismal effects that drugs cause in humans have been attributed to a number of factors such as the interaction with additional target proteins, pathway context, drug–drug interactions, different dosage levels, drug metabolization and aggregation or irreversible target binding of the drug [82–86]. Despite these multiple influencing factors, it is becoming evident that prediction methods based on combining phenotypic information with known molecular activities should be able to generate many novel drug–target relationships.

Furthermore, it seems feasible to exploit the growing amount of data we have reviewed here to considerably enhance our understanding of the various molecular mechanisms underlying the complex effects of existing drugs. This untangling of the various factors influencing drug effects will probably even enable us one day to predict cellular and phenotypic effect of novel drugs.

Acknowledgements: The authors are grateful to Konrad Förstner for discussions. We thank Paul Hawkins and Bob Tolbert from OpenEye Scientific Software for providing the shape comparisons of chemical structures. This work was funded by the BMBF QuantPro Grant (0313831D).

References

[1] Wheeler, D.L. et al. (2007) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 35, D5–D12.
[2] Rognan, D. (2007) Chemogenomic approaches to rational drug design. Br. J. Pharmacol. 152, 38–52.
[3] Chong, C.R. and Sullivan Jr., D.J. (2007) New uses for old drugs. Nature 448, 645–646.
[4] Clayton, T.A. et al. (2006) Pharmaco-metabonomic phenotyping and personalized drug treatment. Nature 440, 1073–1077.
[5] Nicholson, J.K. (2006) Global systems biology, personalized medicine and molecular epidemiology. Mol. Syst. Biol. 2, 52.
[6] Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R. and Gordon, J.I. (2007) The human microbiome project. Nature 449, 804–810.
[7] Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R. and Gordon, J.I. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 444, 1027–1031.
[8] UniProt Consortium (2007) The Universal Protein Resource (UniProt). Nucleic Acids Res. 35, D193–D197.
[9] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. Nucleic Acids Res. 28, 235–242.
[10] Brooksbank, C., Cameron, G. and Thornton, J. (2005) The European Bioinformatics Institute's data resources: towards systems biology. Nucleic Acids Res. 33, D46–D53.
[11] Chen, J.H., Linstead, E., Swamidass, S.J., Wang, D. and Baldi, P. (2007) ChemDB update – full-text search and virtual chemical space. Bioinformatics 23, 2348–2351.
[12] Irwin, J.J. and Shoichet, B.K. (2005) ZINC – a free database of commercially available compounds for virtual screening. J. Chem. Inf. Model. 45, 177–182.
[13] Breitkreutz, B.J., Stark, C. and Tyers, M. (2003) The GRID: the General Repository for Interaction Datasets. Genome Biol. 4, R23.
[14] Kerrien, S. et al. (2007) IntAct – open source resource for molecular interaction data. Nucleic Acids Res. 35, D561–D565.
[15] 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.
[16] Hermjakob, H. et al. (2004) The HUPO PSI's molecular interaction format – a community standard for the representation of protein interaction data. Nat. Biotechnol. 22, 177–183.
[17] Gavin, A.C. et al. (2006) Proteome survey reveals modularity of the yeast cell machinery. Nature 440, 631–636.
[18] Krogan, N.J. et al. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. Nature 440, 637–643.
[19] Lehner, B., Crombie, C., Tischler, J., Fortunato, A. and Fraser, A.G. (2006) Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. Nat. Genet. 38, 896–903.
[20] Tong, A.H. et al. (2004) Global mapping of the yeast genetic interaction network. Science 303, 808–813.
[21] Kuhn, M., von Mering, C., Campillos, M., Jensen, L.J. and Bork, P. (2008) STITCH: interaction networks of chemicals and proteins. Nucleic Acids Res. 36, D684–D688.
[22] Bork, P., Jensen, L.J., von Mering, C., Ramani, A.K., Lee, I. and Marcotte, E.M. (2004) Protein interaction networks from yeast to human. Curr. Opin. Struct. Biol. 14, 292–299.
[23] Seiler, K.P. et al. (2007) ChemBank: a small-molecule screening and cheminformatics resource database. Nucleic Acids Res. 18, 18.
[24] Chen, X., Ji, Z.L. and Chen, Y.Z. (2002) TTD: Therapeutic Target Database. Nucleic Acids Res. 30, 412–415.
[25] Wishart, D.S., Knox, C., Guo, A.C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z. and Woolsey, J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 34, D668–D672.
[26] Günther, S. et al. (2008) SuperTarget and Matador: resources for exploring drug–target relationships. Nucleic Acids Res. 36, D919–D922.
[27] Roth, B.L., Kroeze, W.K., Patel, S. and lopez, E. (2000) The Multiplicity of Serotonin Receptors: uselessly diverse molecules or an embarrasment of riches? Neuroscientist 6, 252–262.
[28] Liu, T., Lin, Y., Wen, X., Jorissen, R.N. and Gilson, M.K. (2007) BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. Nucleic Acids Res. 35, D198–D201.
[29] Roth, B.L., Baner, K., Westkaemper, R., Siebert, D., Rice, K.C., Steinberg, S., Ernsberger, P. and Rothman, R.B. (2002) Salvino-

rin A: a potent naturally occurring nonnitrogenous kappa opioid selective agonist. Proc. Natl. Acad. Sci. USA 99, 11934–11939.
[30] Edgar, R., Domrachev, M. and Lash, A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 30, 207–210.
[31] Parkinson, H. et al. (2007) ArrayExpress – a public database of microarray experiments and gene expression profiles. Nucleic Acids Res. 35, D747–D750.
[32] Vastrik, I. et al. (2007) Reactome: a knowledge base of biologic pathways and processes. Genome Biol. 8, R39.
[33] Kanehisa, M. et al. (2006) From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res. 34, D354–D357.
[34] Caspi, R. et al. (2006) MetaCyc: a multiorganism database of metabolic pathways and enzymes. Nucleic Acids Res. 34, D511–D516.
[35] Apic, G., Ignjatovic, T., Boyer, S. and Russell, R.B. (2005) Illuminating drug discovery with biological pathways. FEBS Lett. 579, 1872–1877.
[36] Rajasethupathy, P., Vayttaden, S.J. and Bhalla, U.S. (2005) Systems modeling: a pathway to drug discovery. Curr. Opin. Chem. Biol. 9, 400–406.
[37] Breitkreutz, B.J. et al. (2008) The BioGRID Interaction Database: 2008 update. Nucleic Acids Res. 36, D637–D640.
[38] Overbeek, R. et al. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. 33, 5691–5702 (Print 2005).
[39] Suderman, M. and Hallett, M. (2007) Tools for visually exploring biological networks. Bioinformatics 23, 2651–2659.
[40] von Mering, C., Jensen, L.J., Kuhn, M., Chaffron, S., Doerks, T., Kruger, B., Snel, B. and Bork, P. (2007) STRING 7 – recent developments in the integration and prediction of protein interactions. Nucleic Acids Res. 35, D358–D362.
[41] Cline, M.S. et al. (2007) Integration of biological networks and gene expression data using Cytoscape. Nat. Protoc. 2, 2366–2382.
[42] Shannon, P. et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504.
[43] Hopkins, A.L., Mason, J.S. and Overington, J.P. (2006) Can we rationally design promiscuous drugs? Curr. Opin. Struct. Biol. 16, 127–136.
[44] Lage, K. et al. (2007) A human phenome–interactome network of protein complexes implicated in genetic disorders. Nat. Biotechnol. 25, 309–316.
[45] Yildirim, M.A., Goh, K.I., Cusick, M.E., Barabasi, A.L. and Vidal, M. (2007) Drug–target network. Nat. Biotechnol. 25, 1119–1126.
[46] Paolini, G.V., Shapland, R.H., van Hoorn, W.P., Mason, J.S. and Hopkins, A.L. (2006) Global mapping of pharmacological space. Nat. Biotechnol. 24, 805–815.
[47] Stelzl, U. et al. (2005) A human protein–protein interaction network: a resource for annotating the proteome. Cell 122, 957–968.
[48] Rual, J.F. et al. (2005) Towards a proteome-scale map of the human protein–protein interaction network. Nature 437, 1173–1178.
[49] Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A. and McKusick, V.A. (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 33, D514–D517.
[50] Willett, P., Barnard, J. and Downs, G. (1998) Chemical similarity searching. J. Chem. Inform. Comp. Sci. 38, 983–996.
[51] Lemmen, C. and Lengauer, T. (2000) Computational methods for the structural alignment of molecules. J. Comp. Aided Mol. Des. 14, 215–232.
[52] Nettles, J.H., Jenkins, J.L., Bender, A., Deng, Z., Davies, J.W. and Glick, M. (2006) Bridging chemical and biological space: "target fishing" using 2D and 3D molecular descriptors. J. Med. Chem. 49, 6802–6810.
[53] Thimm, M., Goede, A., Hougardy, S. and Preissner, R. (2004) Comparison of 2D similarity and 3D superposition. Application to searching a conformational drug database. J. Chem. Inform. Comp. Sci. 44, 1816–1822.
[54] Matter, H. (1997) Selecting optimally diverse compounds from structure databases: a validation study of two-dimensional and three-dimensional molecular descriptors. J. Med. Chem. 40, 1219–1229.
[55] Martin, Y.C., Kofron, J.L. and Traphagen, L.M. (2002) Do structurally similar molecules have similar biological activity? J. Med. Chem. 45, 4350–4358.
[56] Keiser, M.J., Roth, B.L., Armbruster, B.N., Ernsberger, P., Irwin, J.J. and Shoichet, B.K. (2007) Relating protein pharmacology by ligand chemistry. Nat. Biotechnol. 25, 197–206.
[57] Bender, A. et al. (2007) Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure. ChemMedChem 2, 861–873.
[58] Xia, X., Maliski, E.G., Gallant, P. and Rogers, D. (2004) Classification of kinase inhibitors using a Bayesian model. J. Med. Chem. 47, 4463–4470.
[59] Schuffenhauer, A., Floersheim, P., Acklin, P. and Jacoby, E. (2003) Similarity metrics for ligands reflecting the similarity of the target proteins. J. Chem. Inform. Comp. Sci. 43, 391–405.
[60] Mitchell, J.B. (2001) The relationship between the sequence identities of alpha helical proteins in the PDB and the molecular similarities of their ligands. J. Chem. Inform. Comp. Sci. 41, 1617–1622.
[61] Fedorov, O. et al. (2007) A systematic interaction map of validated kinase inhibitors with Ser/Thr kinases. Proc. Natl. Acad. Sci. USA 104, 20523–20528.
[62] Fabian, M.A. et al. (2005) A small molecule–kinase interaction map for clinical kinase inhibitors. Nat. Biotechnol. 23, 329–336.
[63] Chen, J., Zhang, X. and Fernandez, A. (2007) Molecular basis for specificity in the druggable kinome: sequence-based analysis. Bioinformatics 23, 563–572.
[64] Laird, E.R. and Blake, J.F. (2004) Structure-based generation of viable leads from small combinatorial libraries. Curr. Opin. Drug Discov. Dev. 7, 354–359.
[65] Shoichet, B.K., McGovern, S.L., Wei, B. and Irwin, J.J. (2002) Lead discovery using molecular docking. Curr. Opin. Chem. Biol. 6, 439–446.
[66] Bisson, W.H. et al. (2007) Discovery of antiandrogen activity of nonsteroidal scaffolds of marketed drugs. Proc. Natl. Acad. Sci. USA. 104, 11927–11932.
[67] Perez-Iratxeta, C., Bork, P. and Andrade, M.A. (2002) Association of genes to genetically inherited diseases using data mining. Nat. Genet. 31, 316–319.
[68] Peterson, R.T., Shaw, S.Y., Peterson, T.A., Milan, D.J., Zhong, T.P., Schreiber, S.L., MacRae, C.A. and Fishman, M.C. (2004) Chemical suppression of a genetic mutation in a zebrafish model of aortic coarctation. Nat. Biotechnol. 22, 595–599.
[69] Zon, L.I. and Peterson, R.T. (2005) In vivo drug discovery in the zebrafish. Nat. Rev. Drug Discov. 4, 35–44.
[70] Stern, H.M. and Zon, L.I. (2003) Cancer genetics and drug discovery in the zebrafish. Nat. Rev. Cancer 3, 533–539.
[71] Hughes, T.R. et al. (2000) Functional discovery via a compendium of expression profiles. Cell 102, 109–126.
[72] Weinstein, J.N. et al. (1997) An information-intensive approach to the molecular pharmacology of cancer. Science 275, 343–349.
[73] Waring, J.F. et al. (2001) Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. Toxicol. Appl. Pharmacol. 175, 28–42.
[74] Lamb, J. et al. (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science 313, 1929–1935.
[75] Covell, D.G., Wallqvist, A., Huang, R., Thanki, N., Rabow, A.A. and Lu, X.J. (2005) Linking tumor cell cytotoxicity to mechanism of drug action: an integrated analysis of gene expression, small-molecule screening and structural databases. Proteins 59, 403–433.
[76] Lum, P.Y. et al. (2004) Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes. Cell 116, 121–137.
[77] Brown, J.A. et al. (2006) Global analysis of gene function in yeast by quantitative phenotypic profiling. Mol. Syst. Biol. 2, 2006.0001.
[78] Lee, W. et al. (2005) Genome-wide requirements for resistance to functionally distinct DNA-damaging agents. PLoS Genet. 1, e24.
[79] Parsons, A.B. et al. (2006) Exploring the mode-of-action of bioactive compounds by chemical–genetic profiling in yeast. Cell 126, 611–625.

[80] Fliri, A.F., Loging, W.T., Thadeio, P.F. and Volkmann, R.A. (2005) Analysis of drug-induced effect patterns to link structure and side effects of medicines. Nat. Chem. Biol. 1, 389–397.
[81] Zambrowicz, B.P. and Sands, A.T. (2003) Knockouts model the 100 best-selling drugs – will they model the next 100? Nat. Rev. Drug Discov. 2, 38–51.
[82] Liebler, D.C. and Guengerich, F.P. (2005) Elucidating mechanisms of drug-induced toxicity. Nat. Rev. Drug Discov. 4, 410–420.
[83] Demoly, P. and Hillaire-Buys, D. (2004) Classification and epidemiology of hypersensitivity drug reactions. Immunol. Allerg. Clin. N. Am. 24, 345–356, v.
[84] Evans, D.C., Watt, A.P., Nicoll-Griffith, D.A. and Baillie, T.A. (2004) Drug–protein adducts: an industry perspective on minimizing the potential for drug bioactivation in drug discovery and development. Chem. Res. Toxicol. 17, 3–16.
[85] Honig, P.K., Woosley, R.L., Zamani, K., Conner, D.P. and Cantilena Jr., L.R. (1992) Changes in the pharmacokinetics and electrocardiographic pharmacodynamics of terfenadine with concomitant administration of erythromycin. Clin. Pharmacol. Ther. 52, 231–238.
[86] Waring, J.F. and Anderson, M.G. (2005) Idiosyncratic toxicity: mechanistic insights gained from analysis of prior compounds. Curr. Opin. Drug Discov. Dev. 8, 59–65.
[87] Steinbeck, C., Hoppe, C., Kuhn, S., Floris, M., Guha, R. and Willighagen, E.L. (2006) Recent developments of the chemistry development kit (CDK) – an open-source java library for chemo- and bioinformatics. Curr. Pharm. Des. 12, 2111–2120.
[88] Rush 3rd, T.S., Grant, J.A., Mosyak, L. and Nicholls, A. (2005) A shape-based 3-D scaffold hopping method and its application to a bacterial protein–protein interaction. J. Med. Chem. 48, 1489–1495.
[89] Rognes, T. and Seeberg, E. (2000) Six-fold speed-up of Smith–Waterman sequence database searches using parallel processing on common microprocessors. Bioinformatics 16, 699–706.