ELSEVIER

Drug profiling: knowing where it hits

Alejandro Merino[1], Agnieszka K. Bronowska[2], David B. Jackson[3] and Dolores J. Cahill[1,*]

[1] Conway Institute of Biomolecular & Biomedical Research, School of Medicine and Medical Science, University College Dublin, Belfield, Dublin 4, Ireland
[2] EML Research gGmbH, Villa Bosch, Schloss-Wolfsbrunnerweg 33, D-69118 Heidelberg, Germany
[3] LIFE Biosystems GmbH, Poststrasse 34, D-69115 Heidelberg, Germany

Off-target hits of drugs can lead to serious adverse effects or, conversely, to unforeseen alternative medical utility. Selectivity profiling against large panels of potential targets is essential for the drug discovery process to minimize attrition and maximize therapeutic utility. Lately, it has become apparent that drug promiscuity (polypharmacology) goes well beyond target families; therefore, lowering the profiling costs and expanding the coverage of targets is an industry-wide challenge to improve predictions. Here, we review current and promising drug profiling alternatives and commercial solutions in these exciting emerging fields.

Introduction

The old notion that one drug acts on a single protein target has become rapidly obsolete. High- and ultra-high-throughput screenings with a particular validated target have resulted in the identification of many promising compounds. These leads are then channelled through the costly and risky road of prioritization and validation, which, in turn, becomes the 'do or die' dilemma in the pharmaceutical industry. Currently, it is estimated that bringing a drug to market takes from 7 to 12 years and can cost between 500 and 2000 million USD [1,2]. Typically, only one out of 8000 compounds tested in animals reaches human testing, and only one of five compounds reaching clinical trials is eventually approved [3]. Despite this severe scrutiny, more than 120 000 people die every year in the UK and the USA because of side-effects of prescription medicines that now represent the fourth biggest killer in the western world [4]. Several marketed drugs are being withdrawn as a result of serious side-effects, with major consequences not only in patients but also on the entire pharmaceutical industry. Alternative approaches that would lower attrition rates and minimize the risk of new chemical entities are being implemented or eagerly pursued.

This review focuses on the role of drug profiling in the drug discovery process and the emerging *in vitro* approaches and advanced computational methods that the field has adopted recently.

Corresponding author. Cahill, D.J. (dolores.cahill@ucd.ie)

From druggable proteome to druggable systeome

Small compounds normally bind tighter to defined protein pockets that have evolved to bind physiological substrates or ligands. At the end of the past century, it was postulated that the number of druggable targets for commercialized drugs was approximately 500 [5]. Currently, the druggable human proteome for small molecules is estimated at approximately 3000 targets [6]. Approximately 97% of all drugs known today target GPCRs, ion channels, transporters, nuclear receptors and enzymes, most of which have been well characterized by the scientific community.

Recent literature suggests that the druggable human proteome comprises proteins possessing a causal association with disease and an inherent affinity for small-molecule ligands. Although such definition continues to provide a useful principle in target finding and drug discovery pursuits, it is biased towards the 'Ehrlichian' view of therapeutic intervention, with its 'one drug, single target, lone indication' perspective [7]. With the holistic characterization of disease systems now a driving focus of the biomedical research community (e.g. The Cancer Genome Anatomy Project), emerging evidence depicts disease pathways as complex modularized assemblies, constructed and controlled with highly robust design. Not only does this provide some hint as to why many common

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2010.06.006

diseases, such as cancer and diabetes, are multi-factorial in nature but also it enables us to openly question the global applicability of the single-target approach. After all, if a disease is multi-factorial in nature, then it seems logical to extend this idea to the discovery of disease-modifying therapies involving two or more targets. Thus, many proteins considered 'non-druggable' a decade ago might, in fact, achieve druggable status through combined modulation with other proteins in the disease system. From this perspective, one can easily argue that the vast majority of ligand-binding proteins within the human proteome are indeed druggable.

The concept of proteome druggability can today be extended to encompass our growing appreciation of disease system composition and architecture. In fact, as we begin to decipher the cell-type-specific characteristics of a given pathway, so we can start to address the issue of druggability from a more system-orientated perspective. Herein lies the concept of the 'druggable systeome', with the term 'systeome' here implying the content and connectivity of a proteome at the cell-type-specific level. Disease pathways, therefore, represent structured dynamic subsets of a systeome that can, in principle, be compared among cell types to reveal the key molecular features important to drug response. As reviewed previously [8], the NCI60 cancer cell lines are particularly instructive in this regard. Representing nine different human tumours, the NCI60 cancer cell lines have been tested against thousands of compounds, the results of which are publicly available at the Developmental Therapeutic Program website at the National Cancer Institute (http://dtp.nci.nih.gov). Importantly, the molecular composition of the NCI60 cells has also been characterized extensively using a variety of 'omics' technologies. Such molecular details can now be used to construct cell-type-specific models of individual pathways, which can then be classified according to the degree of responsiveness to pathway-targeted therapies. Inter-model comparisons can then be used to reveal the key proteins involved in drug resistance and/or response. Thus, by becoming more specific about the nature of the druggable components within a disease pathway, we can potentially focus drug development efforts on areas of chemistry space that target key mechanistic components of the system.

Considering druggability at this level might have many important advantages. For example, the issue of functional redundancy among system and/or pathway components is an important one and presumably contributes to the lack of efficacy seen in many monotherapy candidates. Using alternative routes for information transfer, signalling pathways are assumed to bypass the effects of antagonists at specific target nodes – a phenomenon known as functional redundancy. Clearly, a multi-nodal therapeutic strategy is desirable in such instances, providing further rational impetus for the application of combinatorial approaches. Although this is best achieved through development of independent targeted therapies, it is not unthinkable to develop a single small molecule that targets multiple important nodes within the pathway (Figure 1). In fact, given the promiscuity of all kinase inhibitors developed to date, such combinatorial effects at the level of the system might be important for achieving clinical effect in robust human disease systems. This is but one reason to emphasize the important conceptual advantages of a druggable systeome view as opposed to that of a druggable proteome. Nevertheless, both concepts remain valid and equally tractable to the technologies described below.

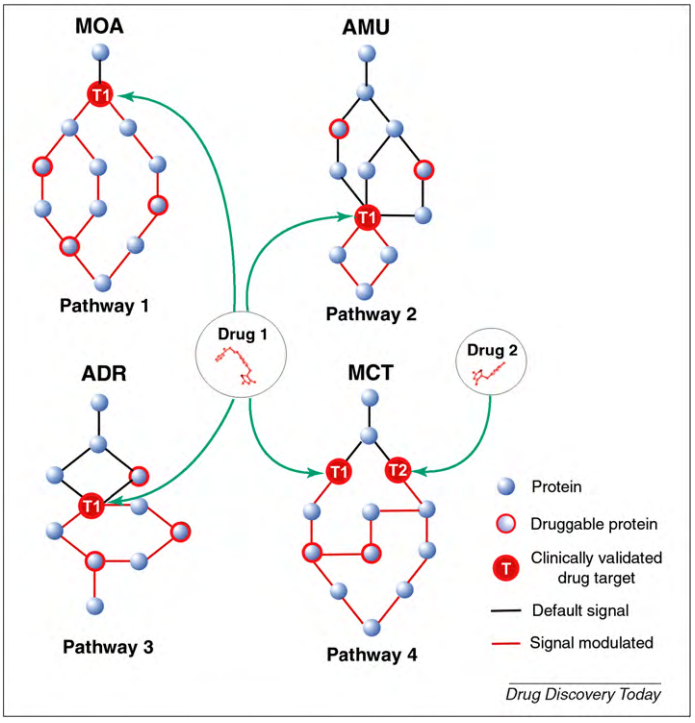

FIGURE 1

Depicting the system-dependent effects of drug target biology. Under standard conditions, a drug modulates a target to achieve beneficial effects on system output; the so-called Mechanism of Action (MOA). However, the occurrence of this target in a different biological context can lead to opportunities for alternative medical use (AMU), and effects on other systems may cause serious adverse drug reactions (ADRs). Finally, in some contexts, modulation of target may not be sufficient to achieve a therapeutic effect, such as when the target is located at a point of functional redundancy. This raises the possibility of targeting other druggable proteins within the system to negate the alternative modes of pathway activity; so-called 'multi-component therapies' (MCTs).

Mechanism of action and toxicity

Despite the broad scientific knowledge around human drug targets, today's pharmacopoeia still includes many drugs that are being prescribed with unknown Mechanism of Action (MOA). Numerous drugs have been approved and marketed based on positive animal model data without a clear understanding of what the drug does at the molecular level. The very reason why some toxicity studies performed in animal models do not reflect what happens in patients might be explained, in part, by the small conservative changes in the drug-binding pockets of targets from different organisms. Thalidomide is a good example of a drug with undefined molecular targets that was thought to be extremely safe in animals. After extensive toxicological tests in animals, thalidomide was considered 'an almost uniquely safe compound' [9]. Nevertheless, in the late 1950s and early 1960s, thalidomide caused an estimated 10 000 birth defects and thousands of fetal deaths worldwide (see 'The tragedy of thalidomide and the failure of animal testing' at http://www.femfatalities.com/thalidomide.asp). The Food and Drug Administration (FDA) based thalidomide's refusal on placental transport, which prevented cases in the USA. This case resulted in the establishment of a more rigorous drug approval process from the early 1960s. Despite tighter regulations, many drugs have been withdrawn in recent years [10]. Vioxx (rofecoxib), another apparently safe drug in several animal

models, including monkeys, had to be withdrawn in 2004 because of the associated risks of blood clots after long-term use that killed 140 000 people worldwide (see 'Vioxx recalled' at http://www.drugrecalls.com/vioxx.html). Vioxx was selected as a specific inhibitor of the catalytic domain of Cox-2. Despite the fact that the amino acid sequence homology of this domain is 92% compared to rabbit, 88% to mouse and 87% to rat, a considerable impact in the K_{on} and/or K_{off} of compounds can be observed. Single amino acid changes can sometimes render a drug completely ineffective [11]. Moreover, off-target binding including metabolizing enzymes, transporters and nuclear receptors might also occur. Therapeutic effects and unforeseen side-effects, therefore, are not necessarily equivalent between animals and humans. Remarkably, mouse studies with Vioxx have shown evidence for an off-target effect that induces the expression of tissue factor, a primary initiator of blood coagulation that could account for Vioxx failure [12]. In most cases, however, such failures but can be attributed to insufficient human *in vitro*, *in silico* and clinical testing.

Conversely, drug testing in animals has also given false warnings regarding toxicity. A well-documented case is that of Glivec (imatinib). This kinase inhibitor showed moderate to severe liver toxicity in rats and dogs and renal toxicity in monkeys, in addition to many other toxicity warnings. Despite this, imatinib was later proven safe in humans (http://www.medicines.ie/medicine/7563/SPC/Glivec). When tested in tissue culture, specific induction of cell death of chronic myelogenous leukaemia cell lines (expressing p210BCR–ABL) was observed. This result, in combination with other positive preclinical data, tilted the balance to move to clinical trials, resulting in the fastest-approved drug in the history of the FDA. Since then, the beneficial effects of imatinib have been extended to other kinases and other cancer types. Imatinib gives extraordinary results in the treatment of not only chronic myelogenous leukaemia and acute lymphoblastic leukaemia but also gastrointestinal stromal cell tumours and metastatic dermatofibrosarcoma protuberans. These diseases are dependent on the expression and activity of the p210BCR–ABL, p185BCR–ABL, c-kit and platelet-derived growth factor receptor kinases, respectively [11,13]. More recently, its alternative medical use has been extended to small cell lung carcinoma, prostate cancer and glioblastoma. This process is known as drug repositioning [14].

In light of this, perhaps it is not surprising that the FDA is continuously increasing demands for the drug discovery industry to define specific targets and potential off-targets, in addition to adsorption, distribution, metabolism and excretion (ADME) data.

Drug promiscuity: multiple targets for better or worse

Hydrophobic compounds in general have the tendency to interact with some well-known promiscuous proteins: (hERG) potassium channel, pregnane X-receptor, cytochrome P450s, P-glycoprotein and some phase II metabolizing enzymes (conjugation enzymes). This small number of 'nontargets' (undesirable targets) can make a big impact on cardiotoxicity (hERG) and/or the metabolism and distribution of most drugs [15]. Profiling compounds against a panel of these proteins is paramount to minimize risks; however, there are also more specific levels of promiscuity or polypharmacology that could lead to very distinctive positive or negative effects [16]. Vasopeptidase inhibitors have peculiar MOA in this context. Omapatrilat, an effective hypertension drug, is a dual inhibitor of angiotensin-converting enzyme (ACE) and neutral endopeptidase. Its combined effect reduces the formation of angiotensin II, and it enables the accumulation of natriuretic factors, such as bradykinin. In this way, Omapatrilat affects both the vascular tone and renal function, making it superior to more specific ACE inhibitors. Triple inhibitors additionally targeting the endothelin-converting enzyme have also been described [17].

Valproic acid (epilepsy and bipolar disorder, solid tumours, and Alzheimer's) is another example of a small molecule that can inhibit several targets, including GSK3 (kinase), HDAC1 (histone deacetylase) and GABA transaminase prolyl oligopeptidase (Figure 2). HDAC1 is needed for HIV to remain in infected cells.

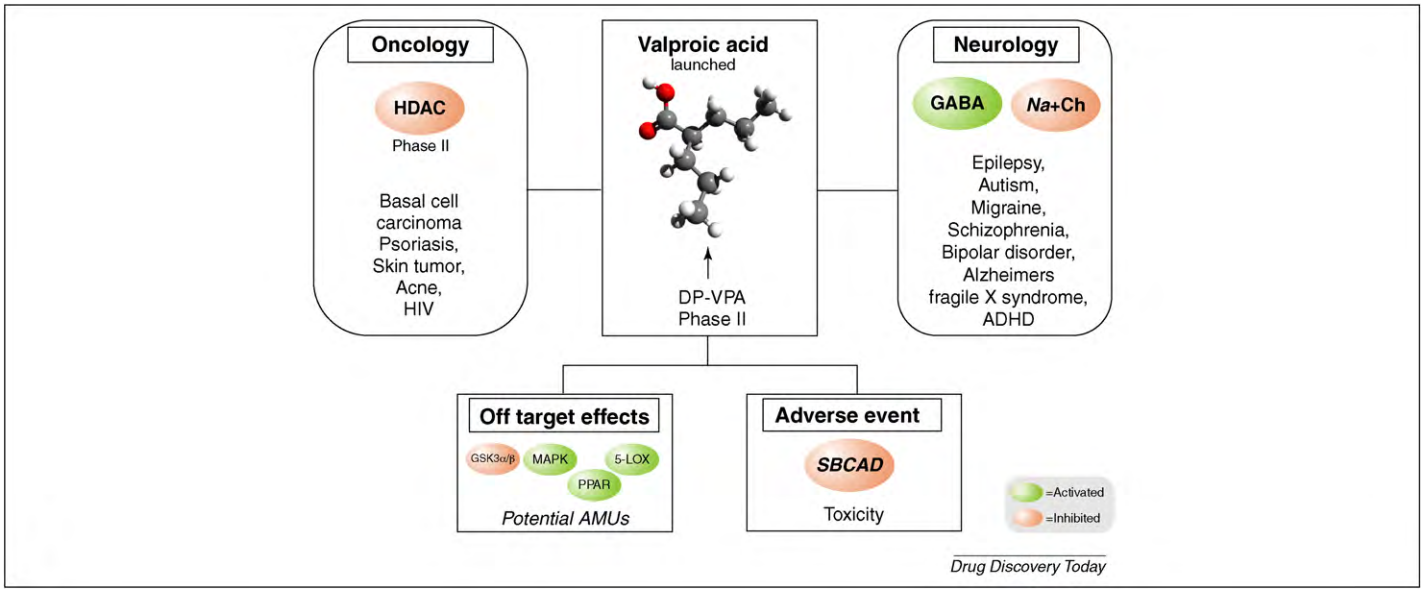

FIGURE 2

Potential targets and therapeutic areas of valproic acid. Abbreviations: AMU, alternative medical use; GABA, GABA levels are increased by inhibition of GABA-metabolizing enzymes; HDAC, histone deacetylase; Na^+Ch, sodium channels; SBCAD, 2-methylbutyryl-CoA dehydrogenase.

In recent years, the treatment of HIV-infected patients with valproic acid in addition to highly active antiretroviral therapy showed a 75% reduction in latent HIV infection, expanding its medical application to AIDS [18]. Further work will have to be performed to define the therapeutic benefits, as well as the adverse effects such as teratogenicity and hepatotoxicity and their potential correlation with its multiple molecular targets.

Several polypharmacological drugs are being revealed by serendipity [18], computational methods [19] or systematic profiling. In a recent review addressing polypharmacology and drug target networks, Vogt and Mestres [20] predict that a drug hits on average 6–13 different targets depending on the profiling method. Small hydrophobic drugs seem to be considerably more promiscuous than large hydrophilic ones. Drugs targeting GPCRs seem to have a wider polypharmacological spectrum than those targeting enzymes, ligand-gated ion channels or nuclear receptors [20]. Given the myriad of possible affinities, it is to be expected that a successful drug is the one for which the target proteins at the top of its affinity list contribute to the desired therapeutic effect, whereas those that at the bottom hold detrimental implications.

When drugs affect many different targets and pathways they are often called 'dirty drugs' [21]. Compounds such as chlorpromazine, dextromethorphan and ibogaine bind to perhaps more than a dozen different types and subclasses of receptors and can cause several physiological and therapeutic effects. Despite the clear therapeutic benefits of some dirty drugs, the regulatory agencies are always looking into alternatives with defined MOA and fewer side-effects. A wider target coverage will be extremely valuable to help ascertain more complete profiles.

Profiling assays

There is a wealth of information on GPCRs, channels, enzymes, MOA and cellular assays in the Assay Guidance Manual on the NIH Chemical Genomics Center website (http://www.ncgc.nih.gov/guidance/manual_toc.html). This website has a comprehensive guide for *in vivo* and *in vitro* assays for compounds that modulate the activity of biological targets. Improvements in functional protein purification and immobilization, detection technologies and assay development are constantly lowering cost and compound demands [22].

In vitro assays

In vitro assays can be based on activity or binding, where both have advantages and disadvantages. The identification of affinities of compounds for the drug binding pockets rather than activity can be very important for enzymes that might need to be activated to screen an inhibitor. Often, inhibitors targeting enzymes that are in the activated state show higher promiscuity among family members. Inhibitors that target enzymes in the inactive configuration, however, seem to be much more selective [11]. This has been very well documented for kinases but could also be extended to other enzymes. Phosphodiesterases have very conserved catalytic domains that are fused to regulatory domains that can keep the enzyme in a close conformation. This high degree of conservation has hindered the development of specific inhibitors without side-effects such as emesis and diarrhoea. The success of Viagra (PDEA5 inhibitor), however, has encouraged research in this area for other conditions such as inflammation and CNS disorders. On the basis of a combination of technologies, including crystallography, modelling, animal models and assay development, a new class of allosteric modulators targeting phosphodiestarase 4D has been described recently [23]. Interestingly, these allosteric PDE4D inhibitors only partially inhibit cAMP hydrolysis with enough selectivity to prevent emesis [23].

Traditionally, 'activity-based' assays require considerably less reagents than their 'direct competition binding' counterparts. It is the amount of compound required at the lead selection stage that prevents rapid and cost-effective prioritization. The EPIC system (http://catalog2.corning.com/Lifesciences/media/pdf/Epic_Apps_CLS_AN_073.pdf), a recent label-free detection technology commercialized by Corning, is expanding biochemical and cell-based assays to 384 microplate formats, enabling the measurement of direct biological interactions in high-throughput. Functional protein microarrays are also being developed to miniaturize formats and expand protein coverage [24].

Cell-based assays

Cytotoxicity assessments cover many different phenotypic changes, including viability and morphological changes (e.g. cell shape, cell membrane integrity, cell motility, and mitochondria function). These events can be measured simultaneously by making use of image-based high content screening systems, which use sophisticated algorithms for automated image processing and analysis [25]. Although these assessments are extremely important, they do not directly define MOA or molecular targets.

Many cell lines have been genetically engineered to mimic aspects of disease biology or to monitor activation of disease-relevant pathways and mechanisms. Using reporter proteins and/or enzymes or specific dyes, one can evaluate with a certain degree of accuracy the effect of a particular compound on a specific protein target. This has been applied extensively for GPCR and ion channels in particular, and new approaches are constantly being developed [26–29].

Affinity assays

Serenex (acquired by Pfizer) started using affinity drug proteomics to define protein targets of ATP-binding proteins. A review of these approaches [30] illustrates different strategies. Several academic and industrial groups have been using immobilized non-specific inhibitors and competing with label-free compounds to evaluate the targets quantitatively using mass spectrometry. So far, the methodologies depend on the affinity and selectivity of the non-specific inhibitors that are used to capture potential drug targets. Although most of the targets evaluated on these attempts contain ATP-binding pockets and promiscuous or abundant proteins, one could conceive the use of compounds to pull down other classes of targets.

Commercial profiling alternatives

Cerep, MDS Pharma Services and Millipore are some of the leading drug profiling companies. Cerep offers profiling for more than 160 different human targets, and MDSPS has approximately 310 enzyme assays and 210 binding assays, mostly for receptors and channels. Cerep's Bioprint is a database that contains the profile of approximately 2500 compounds over 159 assays that can be used

to compare the profile of a new compound [16] (http://www.cerep.fr/Cerep/Users/pages/ProductsServices/bioprintservices.asp).

Millipore now offers functional cell-based assays for 155+ GPCRs, ~50 channels and *in vitro* assays for 350+ kinases and phosphatases. Invitrogen also offers GPCRs, kinases, nuclear receptors and ion channels. Several other companies, such as EMD Biosciences (USA), Caliper Life Sciences (USA), Enzo Lifesciences (USA), Scottish Biomedical (UK), SignalChem (Canada) and Vinci-Biochem (Italy), among others, have provided assays or proteins for the pharmaceutical industry to address increasing demands for drug profiling.

In silico approaches

Computational techniques have greatly influenced drug discovery and their contribution to this process is increasing. Currently, it is estimated that molecular modelling and other computational approaches account for approximately 10% of pharmaceutical R&D expenditure, estimated to rise to approximately 20% by 2016 [3].

Rapid development of computational tools and methodologies applied to drug discovery has been realized because of the increase in computational power, the development of high-throughput techniques, the growing amount of publicly available data and software development. Computational approaches are used to identify drug candidates (hits), to select the most likely candidates for further development (leads), to transform the selected candidates into suitable drugs with reduced side-effects, improving their ADME/Tox properties (lead optimization) and for target assessment (binding promiscuity/polypharmacology.

The most commonly used computational methods for predicting the pharmacological profiles rely on molecular docking and virtual screening. Molecular docking, in principle, predicts the optimal geometry and orientation of drugs upon binding to the cognate molecular target. One of the serious limitations is the need for the high-resolution structure of the protein target, preferably from X-ray crystallography or NMR spectroscopy. Experimental structural data are particularly difficult to obtain for integral membrane proteins, such as GPCRs, which are the targets of approximately 50% of the currently approved drugs [31]. Alternatively, homology modelling, threading and *ab initio* structure predictions in combination with molecular dynamics (MD) and Monte Carlo simulations have shown some utility predicting target structures [32]. An important challenge of docking predictions is the nature of the protein target: its dynamic behaviour, the large number of degrees of freedom and the complexity of its potential energy surface. Other challenges emerge from weak or transient interactions, where the existence of alternative bound orientations further complicates the already complex binding energy landscape [33], and from the failure of correct description of quantum effects, which is particularly evident for the case of halogen bonds [34]. All these reasons combined cause current docking algorithms difficulties in identifying the correct solution from the list of decoys (false positives), and further improvement in the field is needed. Last but not least, despite considerable gains in speed and accuracy [35], screening vast databases of compounds against all the expressed human proteins remains challenging.

Several alternative approaches to molecular docking focus on the chemical structures of the drug, aiming to identify the features (e.g. pharmacophores) responsible for binding to their targets in the absence of a known target structure. Pharmacophore-based methods, for instance, provide a way of establishing a structure–activity relationship for a series of drugs or drug candidates that are known to interact with the target of interest [36]. These approaches are used not only for hit identification (*de novo* design and virtual screening) but also in lead optimization, ADMET predictions and, more recently, polypharmacology predictions [36–38].

Many computational methods for pharmacophore identification have been introduced [39]. As reviewed previously [40], their performance depends on three main factors: addressing the flexibility of the compounds, the alignment technique used to match pharmacophore models and drug molecules, and the definition of pharmacophoric features.

Another group of approaches commonly used in computational drug discovery are based on QSAR, which attempts to quantitatively correlate the structural features or property descriptors of drug molecules with their activity. Among the property descriptors used in QSAR are the parameters for hydrophobicity and solubility, topology, hydrogen-bonding, charge, aromaticity, flexibility, and steric effects. Activity used in QSAR includes *in vitro* and i*n vivo* data. The QSAR approaches are straightforward, much faster than methods requiring explicit structural information (such as molecular docking) and very useful for high-throughput screening of massive databases of compounds [41] or for a combinatorial library design [42]. Their success, however, heavily depends on the quality and accuracy of the input data, right choice of the property descriptors, statistical tools used and the method used for validation of the developed model.

ADME and toxicity predictions

ADME are important considerations in drug discovery and development. Problems with crossing biological barriers (the blood–brain barrier), poor solubility, chemical instability and accumulation can seriously impair the extent to which a compound can be used as a drug, despite its strong and selective binding to its target. The reactions occurring during various stages of the metabolic pathway (e.g. reduction, oxidation, hydrolysis, acetylation and methylation) might cause adverse effects owing to toxicity, catalysis, drug–enzyme interactions and/or drug–drug interactions (allostery, competition, and so on). Drug toxicity itself is an increasing problem. Toxicity has nearly doubled as a cause of drug attrition from 1991 to 2000 [43].

European policy for the evaluation of chemicals strongly advocates the use of computational methods and *ex vivo* assays of toxicity evaluation to minimize bioethical issues of animal testing and save resources [44] (http://www.safermedicines.org).

Recent reviews [3,45] suggest that the most frequently used computational method for predictive toxicology is QSAR combined with statistical approaches. The most serious limitation of use of QSAR methods for toxicity predictions is the growing demand for more toxicity data, both to produce and to validate the models. Two databases of toxicological and clinical endpoints have been created by the FDA (http://www.fda.gov/AboutFDA/CentersOffices/CDER/ucm092150) and ToxNet, a collection of toxicology resources (http://toxnet.nlm.nih.gov). The FDA, in collaboration with software companies, is applying QSAR methods

and data mining techniques for predictive toxicology [3]. Despite the success of the QSAR predictions applied to ADME and toxicology studies, however, concerns still remain about using them for new drugs [46].

Another approach used for computational toxicity predictions relies on the use of so-called 'expert systems' [45]. The computer stores and provides on demand the required information based on the curator's (expert's) input. The main disadvantage of such methodology is the poor sensitivity and, hence, high probability of missing some side-effects [45,47] or of misinterpretation of the data [47].

An alternative is to use structure-based approaches for drug metabolism and toxicity predictions. Currently, there is no single fully integrated structure-based tool able to simulate all the possible processes of drug metabolism [48]; however, a combination of molecular docking, MD simulations and QM calculations can be used successfully for metabolism predictions [48]. The limitations of docking were discussed in the previous section. The main limitation of using MD simulations and free-energy perturbation methods is the need for obtaining long trajectories, which can be very time-consuming. Another limiting factor is the computational cost of performing highly accurate QM calculations, although the recent development of fast and reliable QM methods based on semi-empirical Hamiltonians [49,50] provides an avenue for wider implementation of these methods in predictive toxicology.

Predicting polypharmacology/promiscuity

The computational predictions of drug–target binding (whether target- or ligand-orientated) focus on specificity and strength of interactions between these two species. The lack of specificity resulting from interactions of drugs with off-target proteins has been regarded as 'promiscuity', responsible for undesired side-effects [51]. In many cases, however, such lack of specificity is crucial for the pharmacological profile of the considered drug. Thus, an understanding of the mechanisms governing drug–target interactions and the ability to correctly predict the affinity to multiple targets might explain why some drugs, targeting the same protein, can vary in their side-effects and why certain drugs work better or worse than expected (Figure 3).

As reviewed [52], one way to address the problem of polypharmacology is to compare the drug targets by the similarity of their sequence and/or structure. Another group of tools used for predicting polypharmacology are based on machine-learning methods [51,53,54]. Their aim is to identify binding motifs, which might establish relationships between drugs binding to one target with those binding to another. Typically, these approaches are based on statistical methods, such as Bayesian analyses [55,56].

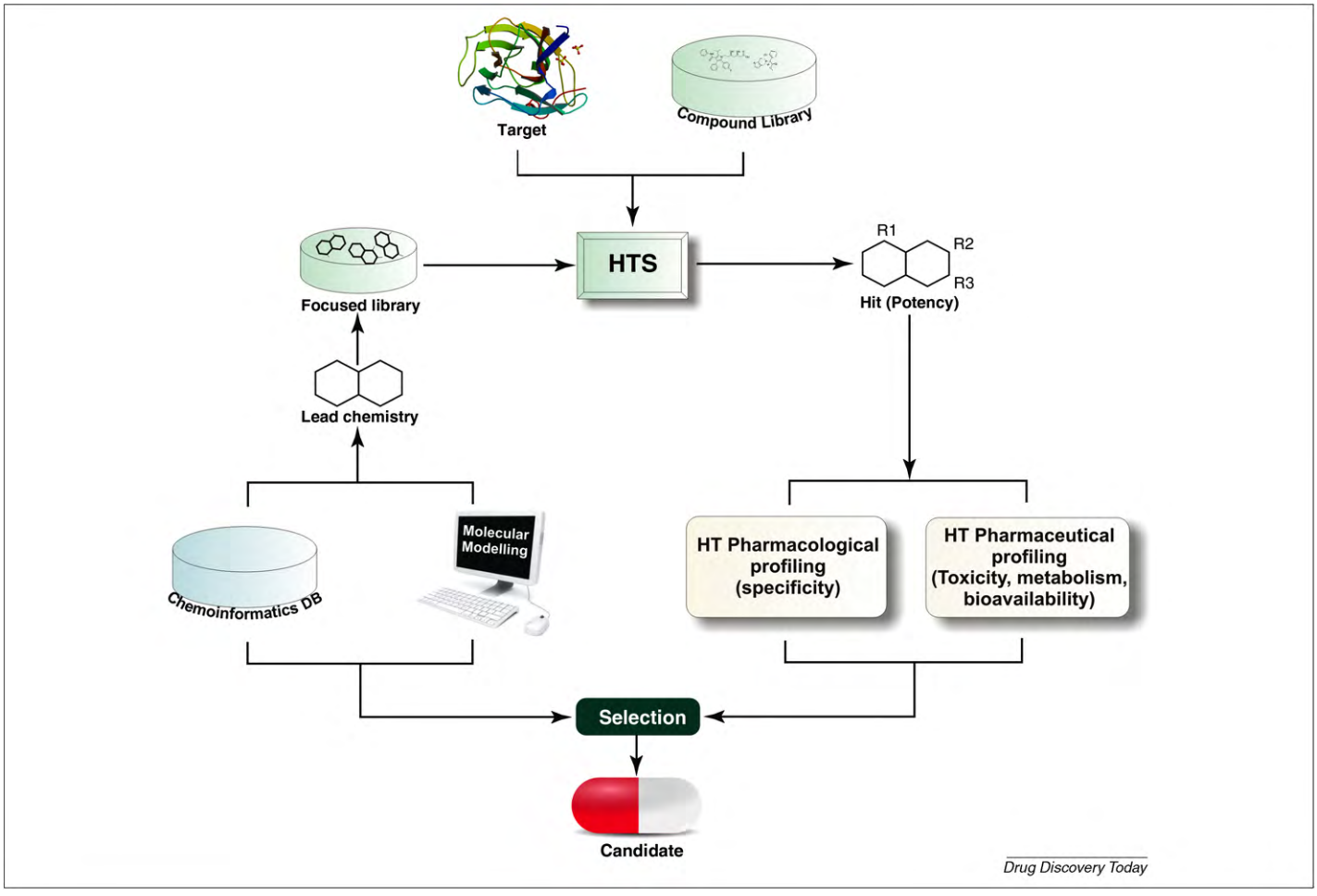

FIGURE 3

Schematic representation of the drug discovery process and the role of pharmacological and pharmaceutical profiling on the selection and prioritization of compounds. HT, high throughput; HTS, high-throughput screening.

Another approach aims to generate predictions of binding promiscuity using drug-based similarities. The data on such similarities can be found in several public databases, such as PubChem [57], TTD [58], ChEMBL (http://www.ebi.ac.uk/chembl/), ChEBI [59], SePreSA [35] and KEGG [60]. These resources provide nomenclature, structure and/or physical properties of small-molecule drugs and, in some cases, their molecular targets. Another important resource, the DrugBank [61] combines detailed drug data with comprehensive information on drug–target interactions and the drug action. One of the very useful components of DrugBank is the information it contains on drug metabolism, drug-metabolizing enzymes and drug-target polymorphisms [61,62]. The comparison of targets based on similarity of drugs that bind to them can be performed using either the BLAST algorithms [63], as in the similarity-ensemble approach [64], or naive Bayesian classifiers [56]. The use of ligand-based similarity to generate predictions of drugs' pharmacological profiles has also been reported [19]. This study used statistics-based cheminformatics approach to predict new off-targets for more than 3000 chemicals. Their approach was proven very successful in both retrospective and new drug-target predictions and could be used for predicting polypharmacology on a much larger scale [19,51].

Accurate predictions of polypharmacology depend on data completeness [65]. The current predictions, such as network topology analysis, obtained using incomplete data, might be misleading and should be taken with caution. The creation and accession of larger databases with more accurate data represents a major challenge for the industry and academic institutions as a whole.

Concluding remarks

The development of *ex vivo* profiling technologies has dramatically changed the way that compounds are tested before entering patients. The ever-increasing panels of assays covering larger portions of the druggable proteome are undoubtedly facilitating our understanding about drug promiscuity to levels that we have never been able to ascertain before. The knowledge acquired from drug profiling is not only considerably lowering the risk of compounds but also shedding new light onto alternative medical utility or drug repositioning. The computational approaches described in this review are already providing helpful insights into drug promiscuity. Systematically validating these predictions will generate more empirical data that, in turn, could evolve into a comprehensive platform for drug profiling. The combination of these technologies together with the classic toxicology and ADME studies will certainly synergize to lower attrition and maximize the utility of new compounds and old drugs.

Acknowledgement

The authors wish to thank David O'Connell for critically reading the manuscript.

References

1 Adams, C.P. and Brantner, V.V. (2006) Estimating the cost of new drug development: is it really 802 million dollars? *Health Aff. (Millwood)* 25, 420–428
2 Shankar, R. *et al.* (2006) LEAN drug development in R&D. *Drug. Discov. Development* 2006, 57–60
3 Kapetanovic, I.M. (2008) Computer-aided drug discovery and development (CADDD): *in silico*-chemico-biological approach. *Chem. Biol. Interact.* 171, 165–176
4 Pirmohamed, M. *et al.* (2004) Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. *BMJ* 329, 15–19
5 Drews, J. and Ryser, S. (1997) The role of innovation in drug development. *Nat. Biotechnol.* 15, 1318–1319
6 Imming, P. *et al.* (2006) Drugs, their targets and the nature and number of drug targets. *Nat. Rev. Drug Discov.* 5, 821–834
7 Copeland, R.A. *et al.* (2006) Drug-target residence time and its implications for lead optimization. *Nat. Rev. Drug Discov.* 5, 730–739
8 Shoemaker, R.H. (2006) The NCI60 human tumour cell line anticancer drug screen. *Nat. Rev. Cancer* 6, 813–823
9 Rowan, A. (1984) *Of Mice, Models and Men: A Critical Evaluation of Animal Research.* State University of New York Press
10 Issa, A.M. *et al.* (2007) Drug withdrawals in the United States: a systematic review of the evidence and analysis of trends. *Curr. Drug Saf.* 2, 177–185
11 Johnson, L.N. (2009) Protein kinase inhibitors: contributions from structure to clinical compounds. *Q. Rev. Biophys.* 42, 1–40
12 Ghosh, M. *et al.* (2007) COX-2 suppresses tissue factor expression via endocannabinoid-directed PPARdelta activation. *J. Exp. Med.* 204, 2053–2061
13 Schindler, T. *et al.* (2000) Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* 289, 1938–1942
14 Tobinick, E.L. (2009) The value of drug repositioning in the current pharmaceutical market. *Drug News Perspect.* 22, 119–125
15 Nobeli, I. *et al.* (2009) Protein promiscuity and its implications for biotechnology. *Nat. Biotechnol.* 27, 157–167
16 Peters, J.U. *et al.* (2009) Pharmacological promiscuity: dependence on compound properties and target specificity in a set of recent Roche compounds. *ChemMedChem* 4, 680–686
17 Daull, P. *et al.* (2007) Towards triple vasopeptidase inhibitors for the treatment of cardiovascular diseases. *J. Cardiovasc. Pharmacol.* 50, 247–256
18 Terbach, N. and Williams, R.S. (2009) Structure-function studies for the panacea, valproic acid. *Biochem. Soc. Trans.* 37, 1126–1132
19 Keiser, M.J. *et al.* (2009) Predicting new molecular targets for known drugs. *Nature* 462, 175–181
20 Vogt, I. and Mestres, J. (2010) Drug-target networks. *Mol. Inf.* 29, 10–14
21 Frantz, S. (2005) Drug discovery: playing dirty. *Nature* 437, 942–943
22 Terstappen, G.C. *et al.* (2007) Target deconvolution strategies in drug discovery. *Nat. Rev. Drug Discov.* 6, 891–903
23 Burgin, A.B. *et al.* (2010) Design of phosphodiesterase 4D (PDE4D) allosteric modulators for enhancing cognition with improved safety. *Nat. Biotechnol.* 28, 63–70
24 Campbell, C.T. and Kim, G. (2007) SPR microscopy and its applications to high-throughput analyses of biomolecular binding events and their kinetics. *Biomaterials* 28, 2380–2392
25 Young, D.W. *et al.* (2008) Integrating high-content screening and ligand-target prediction to identify mechanism of action. *Nat. Chem. Biol.* 4, 59–68
26 Brough, S.J. and Shah, P. (2009) Use of aequorin for G protein-coupled receptor hit identification and compound profiling. *Methods Mol. Biol.* 552, 181–198
27 Lee, P.H. (2009) Label-free optical biosensor: a tool for G protein-coupled receptors pharmacology profiling and inverse agonists identification. *J. Recept. Signal Transduct. Res.* 29, 146–153
28 Haasen, D. *et al.* (2008) Pharmacological profiling of chemokine receptor-directed compounds using high-content screening. *J. Biomol. Screen.* 13, 40–53
29 MacDonald, M.L. *et al.* (2006) Identifying off-target effects and hidden phenotypes of drugs in human cells. *Nat. Chem. Biol.* 2, 329–337
30 Bantscheff, M. *et al.* (2009) Revealing promiscuous drug-target interactions by chemical proteomics. *Drug Discov. Today* 14, 1021–1029
31 Lundstrom, K. (2009) An overview on GPCRs and drug discovery: structure-based drug design and structural biology on GPCRs. *Methods Mol. Biol.* 552, 51–66
32 Ritchie, D.W. (2008) Recent progress and future directions in protein-protein docking. *Curr. Protein Pept. Sci.* 9, 1–15
33 Pons, C. *et al.* (2010) Present and future challenges and limitations in protein-protein docking. *Proteins* 78, 95–108
34 Lu, Y. *et al.* (2009) Halogen bonding-a novel interaction for rational drug design? *J. Med. Chem.* 52, 2854–2862
35 Yang, L. *et al.* (2009) SePreSA: a server for the prediction of populations susceptible to serious adverse drug reactions implementing the methodology of a chemical-protein interactome. *Nucleic Acids Res.* 37 (Web Server issue), W406-412
36 Sun, H. (2008) Pharmacophore-based virtual screening. *Curr. Med. Chem.* 15, 1018–1024

37 Klabunde, T. (2007) Chemogenomic approaches to drug discovery: similar receptors bind similar ligands. *Br. J. Pharmacol.* 152, 5–7

38 Rognan, D. (2007) Chemogenomic approaches to rational drug design. *Br. J. Pharmacol.* 152, 38–52

39 Dror, O. *et al.* (2004) Predicting molecular interactions *in silico*: I. A guide to pharmacophore identification and its applications to drug design. *Curr. Med. Chem.* 11, 71–90

40 Schneidman-Duhovny, D. *et al.* (2008) PharmaGist: a webserver for ligand-based pharmacophore detection. *Nucleic Acids Res.* 36 (Web Server issue), W223-228

41 Gedeck, P. and Lewis, R.A. (2008) Exploiting QSAR models in lead optimization. *Curr. Opin. Drug Discov. Devel.* 11, 569–575

42 Eitner, K. and Koch, U. (2009) From fragment screening to potent binders: strategies for fragment-to-lead evolution. *Mini Rev. Med. Chem.* 9, 956–961

43 Kola, I. and Landis, J. (2004) Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* 3, 711–715

44 Simon-Hettich, B. *et al.* (2006) Use of computer-assisted prediction of toxic effects of chemical substances. *Toxicology* 224, 156–162

45 Merlot, C. (2009) Computational toxicology-a tool for early safety evaluation. *Drug Discov. Today*

46 Bhogal, N. *et al.* (2005) Toxicity testing: creating a revolution based on new technologies. *Trends Biotechnol.* 23, 299–307

47 Muster, W. *et al.* (2008) Computational toxicology in drug development. *Drug Discov. Today* 13, 303–310

48 Sun, H. and Scott, D.O. (2009) Structure-based drug metabolism predictions for drug design. *Chem. Biol. Drug Des.* 75, 3–17

49 Stewart, J.J. (2009) Application of the PM6 method to modeling proteins. *J. Mol. Model.* 15, 765–805

50 Rezac, J. *et al.* (2009) Semiempirical quantum chemical PM6 method augmented by dispersion and H-bonding correction terms reliably describes various types of noncovalent complexes. *J. Chem. Theory Comput.* 5, 1749–1760

51 Hopkins, A.L. (2009) Drug discovery: predicting promiscuity. *Nature* 462, 167–168

52 Schalon, C. *et al.* (2008) A simple and fuzzy method to align and compare druggable ligand-binding sites. *Proteins* 71, 1755–1778

53 Fukunishi, Y. (2009) Structure-based drug screening and ligand-based drug screening with machine learning. *Comb. Chem. High Throughput Screen.* 12, 397–408

54 Park, K. *et al.* (2009) Predicting the multi-modal binding propensity of small molecules: towards an understanding of drug promiscuity. *Mol. Biosyst.* 5, 844–853

55 Bender, A. *et al.* (2006) ''Bayes affinity fingerprints'' improve retrieval rates in virtual screening and define orthogonal bioactivity space: when are multitarget drugs a feasible concept? *J. Chem. Inf. Model.* 46, 2445–2456

56 Klon, A.E. (2009) Bayesian modeling in virtual high throughput screening. *Comb. Chem. High Throughput Screen.* 12, 469–483

57 Wheeler, D.L. *et al.* (2008) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 36 (Database issue), D13–D21

58 Chen, X. *et al.* (2002) TTD: Therapeutic Target Database. *Nucleic Acids Res.* 30, 412–415

59 Brooksbank, C. *et al.* (2005) The European Bioinformatics Institute's data resources: towards systems biology. *Nucleic Acids Res.* 33(Database issue), D46-D53

60 Kanehisa, M. *et al.* (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* 38 (Database issue), D355-D360

61 Wishart, D.S. et al. (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* 36 (Database issue), D901-D906

62 Wishart, D.S. (2008) DrugBank and its relevance to pharmacogenomics. *Pharmacogenomics* 9, 1155–1162

63 Jones, D.T. and Swindells, M.B. (2002) Getting the most from PSI-BLAST. *Trends Biochem. Sci.* 27, 161–164

64 Hert, J. *et al.* (2008) Quantifying the relationships among drug classes. *J. Chem. Inf. Model.* 48, 755–765

65 Mestres, J. *et al.* (2008) Data completeness-the Achilles heel of drug-target networks. *Nat. Biotechnol.* 26, 983–984