British Journal of Pharmacology (2007) 152, 9–20
 0007–1188/07 $30.00
www.brjpharmacol.org

REVIEW

In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling

S Ekins[1,2], J Mestres[3] and B Testa[4]

[1]*ACT LLC, New York, NY, USA;* [2]*Department of Pharmaceutical Sciences, University of Maryland, Baltimore, MD, USA;* [3]*Chemogenomics Laboratory, Research Unit on Biomedical Informatics, Institut Municipal d'Investigació Mèdica and Universitat Pompeu Fabra, Parc de Recerca Biomèdica, Barcelona, Spain and* [4]*University Hospital Centre, Lausanne, Switzerland*

Pharmacology over the past 100 years has had a rich tradition of scientists with the ability to form qualitative or semi-quantitative relations between molecular structure and activity *in cerebro*. To test these hypotheses they have consistently used traditional pharmacology tools such as *in vivo* and *in vitro* models. Increasingly over the last decade however we have seen that computational (*in silico*) methods have been developed and applied to pharmacology hypothesis development and testing. These *in silico* methods include databases, quantitative structure-activity relationships, pharmacophores, homology models and other molecular modeling approaches, machine learning, data mining, network analysis tools and data analysis tools that use a computer. *In silico* methods are primarily used alongside the generation of *in vitro* data both to create the model and to test it. Such models have seen frequent use in the discovery and optimization of novel molecules with affinity to a target, the clarification of absorption, distribution, metabolism, excretion and toxicity properties as well as physicochemical characterization. The aim of this review is to illustrate some of the *in silico* methods for pharmacology that are used in drug discovery. Further applications of these methods to specific targets and their limitations will be discussed in the second accompanying part of this review.
British Journal of Pharmacology (2007) **152,** 9–20; doi:10.1038/sj.bjp.0707305; published online 4 June 2007

Keywords: quantitative structure–activity relationships; homology models; docking; pharmacology; ADME/Tox; *in silico* *in vitro*; drug discovery; computational

Abbreviations: ADME/Tox, absorption, distribution, metabolism, excretion and toxicity; CNS, central nervous system; CoMFA, comparative molecular field analysis; CoMSIA, comparative molecular similarity indices analysis; HTS, high-throughput screening; PD, pharmacodynamic; PK, pharmacokinetic; PRT, purine phosphoribosyltransferase; QSAR, quantitative structure–activity relationship

Introduction

The term '*in silico*' is a modern word usually used to mean experimentation performed by computer and is related to the more commonly known biological terms *in vivo* and *in vitro*. The history of the '*in silico*' term is poorly defined, with several researchers claiming their role in its origination. However, some of the earliest published examples of the word include the use by Sieburg (1990) and Danchin *et al*. (1991). In a more recent book, Danchin (2002) provides a quotation that offers a concise and cogent depiction of the potential of computational tools in chemistry, biology and pharmacology:

> '[…] [I]nformatics is a real aid to discovery when analyzing biological functions […]. […] I was convinced of the potential of the computational approach, which I called *in silico*, to underline its importance as a complement to *in vivo* and *in vitro* experimentation.'

In silico pharmacology (also known as computational therapeutics, computational pharmacology) is a rapidly growing area that globally covers the development of techniques for using software to capture, analyse and integrate biological and medical data from many diverse sources. More specifically, it defines the use of this information in the creation of computational models or simulations that can be used to make predictions, suggest hypotheses, and ultimately provide discoveries or advances in medicine and therapeutics.

Time and again it has been stated that the successful industrial companies are those that manage information as a

Correspondence: Dr S Ekins, ACT LLC, 1 Penn Plaza-36th Floor, New York, NY 10119, USA.
E-mail: ekinssean@yahoo.com, sekins@arnoldllc.com
Received 21 December 2006; revised 27 February 2007; accepted 25 April 2007; published online 4 June 2007

key resource. We could reiterate this for drug discovery, which is a hugely complex information handling and interpretation exercise. With so much information to process, we need to be able to discover the shortcuts or the rules that will point us as quickly as possible to the targets and molecules that are likely to proceed to the clinic then onto the market. Computational or *in silico* methods are helping us to make decisions and simulate virtually every facet of drug discovery and development (Swaan and Ekins, 2005), moving the pharmaceutical industry closer to engineering-based disciplines. For example, we can cherry pick ideas or molecules using virtual screening (Lengauer *et al*., 2004; Shoichet, 2004) as described later.

It has also been suggested that if we are to build on the advances of the human genome, we need to integrate computational and experimental data, with the aim of initiating *in silico pharmacology* linking all data types. This could change the way the pharmaceutical industry discovers drugs using data to enable simulations; however, there may still be significant gaps in our knowledge beyond genes and proteins (Whittaker, 2003). Structure-based methods are broadly used for drug discovery but these are just a beginning, for example in neuropharmacology, it is expected that ligand–receptor interaction kinetic models will need to be integrated with network approaches to understand fully neurological disorders, in general this could be applied more widely to pharmacology (Aradi and Erdi, 2006).

Basically, there are two outcomes when bioactive compounds and biological systems interact (Figure 1) (Testa and Krämer, 2006). Note that 'biological system' is defined here very broadly and includes functional proteins (for example, receptors), monocellular organisms and cells isolated from multicellular organisms, isolated tissues and organs, multicellular organisms and even populations of individuals, be they uni- or multicellular. As for the interactions between a drug (or any xenobiotic) and a biological system, they may be simplified to 'what the compound does to the biosystem' and 'what the biosystem does to the compound.' A drug that acts on a biological system can elicit a pharmacological and/or toxic response, in other words a pharmacodynamic (PD) event.

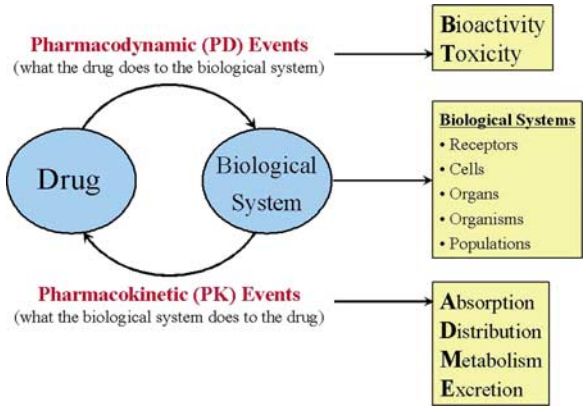


Figure 1 The two basic modes of interaction between xenobiotics and biological systems, namely PD (activity and toxicity) and PK events (ADME) (modified from (Testa and Krämer, 2006) and reproduced with the kind permission of the Verlag Helvetica Chimica Acta in Zurich). ADME, absorption, distribution, metabolism and excretion; PD, pharmacodynamic; PK, pharmacokinetic.

Symmetrically, the biological system acts on the xenobiotic by absorbing, distributing, metabolising and excreting it. These are the pharmacokinetic (PK) events. But one must appreciate that these two aspects of the behaviour of xenobiotics are inextricably interdependent. Absorption, distribution and elimination will obviously have a decisive influence on the intensity and duration of PD effects, whereas biotransformation will generate metabolites that may have distinct PD effects of their own. Conversely, by its own PD effects, a compound may affect the state of the organism (for example, hemodynamic changes and enzyme activities) and hence its capacity to handle xenobiotics. Only a systemic approach as used in PK/PD modelling and in clinical pharmacology is capable of appreciating the global nature of this interdependence. To clarify this discussion, it may be useful to designate as *targets* the various biological components that elicit a PD event following their interaction with a drug or another xenobiotic. Such targets include receptors, ion channels, nucleic acids, anabolic and catabolic enzymes, and so on. Similarly, one can refer to *agents* for the biological components (xenobiotic-metabolising enzymes, transporters, circulating proteins, membranes, and so on) which act on drugs by metabolising, transporting, distributing or excreting them.

History and evolution of *in silico* approaches

Drug design and related disciplines in drug discovery did not wait for the advent of informatics to be born and to grow as sciences. As masterfully summarised by Albert (1971, 1985), the earliest intuitions and insights in structure–activity relations can be traced to the nineteenth century. A relation between activity and a physicochemical property was firmly established by Meyer (1899) and Overton (1901), who proposed a 'Lipoid theory of cellular depression' such that the higher the partition coefficient between a lipid solvent and water, the greater the depressant action. Such papers paved the way for the recognition of lipophilicity and electronic properties as major determinants of PD and PK responses, as best illustrated by the epoch-making and still ongoing work of Corwin Hansch (Hansch and Fujita, 1964; Hansch, 1972), a founding father of drug design.

Other pioneers (for example, Crum Brown and Fraser; reviewed by (Albert, 1971)) saw that chemical structure (that is, the nature and connectivity of atoms in a molecule, in fact the two-dimensional structure (2D) of compounds) also played an essential role in pharmacological activity. The conceptual jump from 2D to three-dimensional (3D) structure owes much to the work of Cushny (1926), whose book summarises a life dedicated to relations between enantiomerism and bioactivity. This vision was expanded in the mid-twentieth century by the discovery of conformational effects on bioactivity (Burgen, 1981).

In parallel with our growing understanding of the concept of molecular structure, a few visionary investigators in the late nineteenth and early twentieth centuries (for example, John Langley, Paul Ehrlich and Alfred Clark; reviewed by (Arïens, 1979; Parascandola, 1980) developed the concept of receptors, namely the targets of drug action. The analogies between receptors and enzymes were outlined by Albert (1971).

The converging lines of progress in chemistry and biology generated a flood of information and knowledge which went beyond the usual capacity of '*in cerebro*' data handling and was a driving force in the emergence and development of computer sciences. Hansch was among the very first in the 1950s to use calculators and statistics to arrive at quantitative relations between structure (in fact, parameters and descriptors) and activity. Such was the birth of quantitative structure–activity relationships (QSARs), followed in the 1980s and 1990s by computer graphics and molecular modelling. However, computer sciences rapidly ceased to be a simple tool in drug discovery and pharmacology and became a major contributor to progress. The chemistry–biology–informatics triad has now evolved into a life of its own and is bringing pharmacology to new heights, as this review will briefly attempt to illustrate.

Quantitative structure–activity relationships

The infancy of *in silico* pharmacology can be established in the early 1960s when quantitative relationships between chemical structure and PD and PK effects in biological systems began to be unveiled by computational means. Since then, the analysis and recognition of QSAR has become an essential component of modern medicinal chemistry and pharmacology. The initial focus was in providing computational estimates for the bioactivity of molecules (Hansch and Fujita, 1964). Accordingly, in a clear break from nomenclature, any attempt being made to establish a connection between chemical structure and a biological effect (that being activity, toxicity, absorption, distribution, metabolism, excretion and toxicity (ADME/Tox) or physico-chemical properties) will be generally referred to in this review as QSAR, for clarity. Therefore, in their broadest sense, QSARs consist of construction of a mathematical model relating a molecular structure to a chemical property or biological effect by means of statistical techniques. This is not an easy task when considering, on the one hand, the possibility that different molecules act by different mechanisms or interact with the receptor in different binding modes leading to the presence of outliers which are unable to fit any QSAR model (Verma and Hansch, 2005) but also, on the other hand, the intrinsic noise associated with both the original data and concrete methodological aspects involved in the construction of a QSAR model (Polanski *et al*., 2006). Ultimately, if a significant correlation is achieved for a set of training molecules for which robust biological data is available, the model can then be used to predict the biological effect for other molecules, although as will be described in the accompanying review there may be some limitations to model applicability that should be considered (Ekins *et al*., 2007). Over the last 40 years, these efforts have generated thousands of QSAR models, many of which have been collected and stored in the C-QSAR database (Hansch *et al*., 2002; Kurup, 2003).

Descriptor-based methods. A key aspect in QSAR is the use of molecular descriptors as numerical representations of chemical structures. The number and type of molecular descriptors is large and varied (Karelson, 2000; Todeschini and Consonni, 2000) and thus procedures to select those that are most relevant for modelling the biological effect of interest are extremely important (Walters and Goldman, 2005). Molecular descriptors are usually classified according to the dimensionality of the chemical representation from which they are computed (Xue and Bajorath, 2000). On this basis, one-dimensional descriptors encode numerically generic properties such as molecular weight, molar refractivity and octanol/water partition coefficient, offering a fair reflection of the size, shape and lipophilicity of molecules. Despite their low dimensionality, some of these descriptors have been associated with the drug-like character of molecules (Lipinski *et al*., 1997) and are thus found often as biologically relevant descriptors in QSAR equations (Hansch *et al*., 2002). On the other hand, 2D descriptors are computed from topological representations of molecules (Gozalbes *et al*., 2002). The models constructed from these descriptors are habitually referred to as 2D-QSAR, a methodology widely established both in predicting physico-chemical properties as well as in providing quantitative estimates of various biological effects (Dudek *et al*., 2006).

In contrast, 3D descriptors are obtained directly from the 3D structure of molecules thus resulting in the so-called 3D-QSAR methods (Akamatsu, 2002). A characteristic of 3D descriptors is their dependence on the molecular conformation being used. This is the reason why many 3D-QSAR methods require the molecules that are aligned before constructing the model (Kubinyi *et al*., 1998), which has in part motivated the emergence of multiple techniques and approaches for aligning molecules, either directly through the flexible superposition of molecules (Lemmen and Lengauer, 2000) or through the docking of compounds in a protein active site, in cases when experimentally determined structural information on the target protein is available (Ortiz *et al*., 1995; Sippl, 2002). Of mention among these alignment-dependent 3D-QSAR methods are the widely established comparative molecular field analysis (Cramer *et al*., 1988) and comparative molecular similarity indices analysis (Klebe, 1998). This limitation motivated the development of alternative alignment-independent 3D-QSAR methods based on the statistical analysis of distributions of surface-based features (Pastor *et al*., 2000; Stiefl and Baumann, 2003).

All these advancements were made with the hope that, compared to 2D-QSAR, 3D-QSAR methods would lead to statistically better models, as the type of descriptors used are in principle more representative of the molecular features that are exposed when interacting with proteins. Unfortunately, experience in a large number and diverse range of applications over the last four decades shows this is not always the case (Perkins *et al*., 2003; Chohan *et al*., 2006). Developing the optimal QSAR model for the chemical property or biological effect of interest, as well as defining its applicability domain within chemical space, is still very much an active area of investigation (Dimitrov *et al*., 2005) as described later (Ekins *et al*., 2007).

Rule-based methods. Statistically relevant QSAR models are usually derived first from training sets composed of a few tens of molecules to be assessed, then in a second stage, on

an external set of molecules. The availability of biological data for an increasing amount of ligands and protein–ligand complexes has allowed the appearance of different types of approaches. These are based on maximally exploiting this information to extract knowledge and derive rules that can then be applied to estimate quantitatively potential biological effects of molecules from structure. A good example is the rule-based methods derived from human expertise on the biotransformation of ligands for its application in the prediction of sites labile to drug metabolism. In these methods, the rule-based algorithm first recognises target sites (that is, functional groups) in the query molecule, then lists all potential metabolic transformations these sites can undergo, and finally prioritises the resulting metabolites based on rules derived from prior knowledge (Kulkarni *et al*., 2005). Existing systems of this type are *MetabolExpert* (Darvas *et al*., 1999), *META* (Klopman and Tu, 1999) and *METEOR* (Langowski and Long, 2002). In particular, the *METEOR* system contains a biotransformation dictionary describing over 300 reaction rules and more than 800 reasoning rules. The metabolic reactions are descriptions of generic reactions rather than simple entries in a reaction database. The reasoning engine contains rules of two types, namely absolute and relative. The rules of absolute reasoning evaluate the likelihood of a biotransformation taking place based on five levels: *Probable*, *Plausible*, *Equivocal*, *Doubted* and *Improbable* (Button *et al*., 2003). The rules of relative reasoning assign priorities to potentially competing reactions (for example, primary alcohols are oxidised in preference to secondary alcohols), equal priority being assigned when no preference is known. The reasoning engine then uses the non-numerical Logic of Argumentation to construct arguments for and against a hypothesis and hence evaluate the likelihood of a specific reaction taking place in the query substrate. The likelihood of biotransformation can also be modified by the reasoning engine according to the general, global relationship between drug metabolism and lipophilicity thanks to a link to an external log P predictor. The reasoning engine uses further rules to avoid a combinatorial explosion of output resulting from unconstrained analyses of query structures. Queries can be analysed at a number of available search levels, such that only reactions of greater likelihood than the chosen threshold are displayed. A recent application of this approach to galantamine showed a full qualitative agreement between *in vivo* experimental results in rats, dogs and humans (Mannens *et al*., 2002) and the *in silico* predicted biotransformations (Testa *et al*., 2005) (Figure 2).

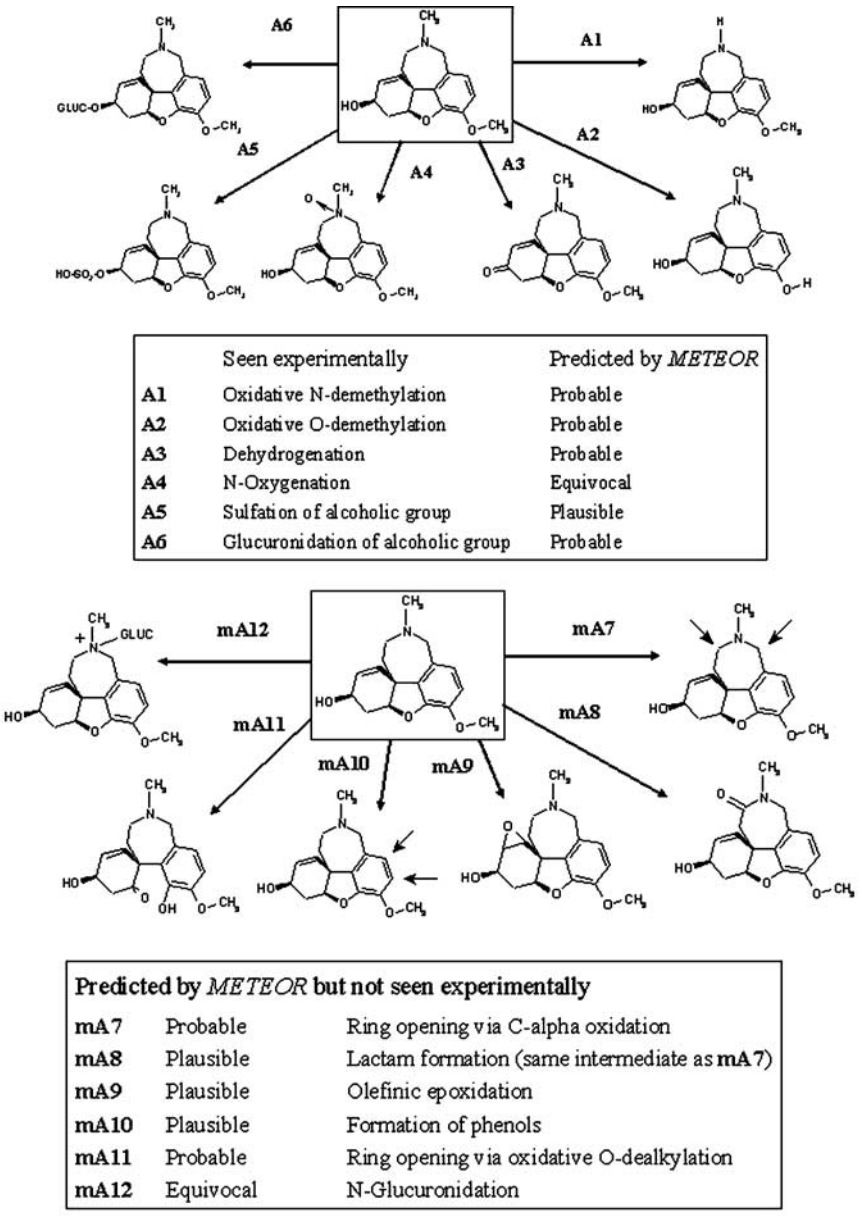
Knowledge-based approaches. In a different field, databases of ligand–protein complexes are being exploited to derive knowledge-based potentials as a means to estimate the free energies of molecular interactions when docking ligands into protein cavities (Gohlke and Klebe, 2002). This knowledge-based approach essentially involves converting interatomic distance contributions found in ligand–protein complexes into pair-potential functions for the different pairs of ligand/protein atom types. An estimation of the free energy of interaction between a ligand and a protein is then obtained by adding the contributions from ligand/protein atom pairs within a certain distance. An attractive feature of this approach is that important, but poorly understood, contributions to ligand–protein binding (such as entropic terms and solvation) are implicitly taken into account. The different potentials differ mainly on the ligand and protein atom types defined, the nature and extent of the experimental set of complexes used, the range of interatomic distances scanned and the width of the distance bins. The relative performance of four knowledge-based potentials (BLEEP-2, DrugScore, PMF and SMoG2001) for estimating the ligand-binding affinities for a set of 77 complexes representative of the same number of proteins was reviewed recently (Fradera and Mestres, 2004). None of the knowledge-based potentials consistently outperformed the rest. The results revealed that present knowledge-based potentials are still far from being universally applicable, current performance being strongly dependent on the type of ligand–protein complexes being analysed. There has been some discussion as to what steps can be taken to improve the protein–ligand potentials to balance speed and accuracy while enabling the efficient use of data and maximising transferability (Shimada, 2006).

Virtual ligand screening

The process of scoring and ranking molecules in large chemical libraries according to their likelihood of having affinity for a certain target, is generally referred to as virtual screening (Oprea and Matter, 2004). In this respect, virtual screening can be regarded as an attempt to extend the concept of QSAR, originally focused on small sets of congeneric compounds, along the chemical dimension defined by existing synthesised molecules as well as plausible synthesisable molecules. The term itself was coined in the late 1990s when computer-based methods reached sufficient maturity to offer an alternative to experimental high-throughput screening (HTS) techniques that were having disappointingly poorer performances and higher costs than originally anticipated (Lahana, 1999). Over the years, the pharmaceutical industry has learnt to accept that virtual screening methods can indeed be an efficient complement to HTS (Stahura and Bajorath, 2004) to the point that they have undoubtedly become an integral part of today's lead generation process (Bajorath, 2002; Bleicher *et al*., 2003).

In contrast to technology-driven HTS, virtual screening is a knowledge-driven approach that requires structural information either on bioactive ligands for the target of interest (ligand-based virtual screening) or on the target itself (target-based virtual screening). Comparative studies have suggested that information about a target obtained from known bioactive ligands is as valuable as knowledge of the target structures for identifying novel bioactive scaffolds through virtual screening (Evers *et al*., 2005a; Zhang and Muegge, 2006). Therefore, the final choice for a method to use will ultimately depend on the type and amount of information available without *a priori* having a large impact on performance.

Ligand-based methods. A diverse range of ligand-based virtual screening methods exist. Their degree of sophistica-



	Seen experimentally	Predicted by *METEOR*
A1	Oxidative N-demethylation	Probable
A2	Oxidative O-demethylation	Probable
A3	Dehydrogenation	Probable
A4	N-Oxygenation	Equivocal
A5	Sulfation of alcoholic group	Plausible
A6	Glucuronidation of alcoholic group	Probable

Predicted by *METEOR* but not seen experimentally		
mA7	Probable	Ring opening via C-alpha oxidation
mA8	Plausible	Lactam formation (same intermediate as **mA7**)
mA9	Plausible	Olefinic epoxidation
mA10	Plausible	Formation of phenols
mA11	Probable	Ring opening via oxidative O-dealkylation
mA12	Equivocal	N-Glucuronidation

Figure 2 Metabolic scheme of galantamine comparing the experimental *in vivo* results in rats, dogs and humans (Mannens *et al.*, 2002) with the predictions of *METEOR* (Testa *et al.*, 2005). (Reproduced with the kind permission of the Verlag Helvetica Chimica Acta in Zurich).

tion, and thus their ultimate computational cost, depend very much on the type of structural information being used (Lengauer *et al.*, 2004). All these methods rely on the central similarity-property principle which states that similar molecules should exhibit similar properties (Johnson and Maggiora, 2006) and thus chemical similarity calculations are at the core of ligand-based virtual screening (Willett *et al.*, 2003). Accordingly, all the molecules in a particular database can be scored relative to the similarity to one or multiple bioactive ligands and then ranked to reflect decreasing probability of being active. These methods generally provide significant enrichments over random selection of molecules in databases. After this procedure, the top scoring molecules can be prioritised for going into experimental testing and thus represents a cost-effective strategy in drug discovery programmes.

Of the different structure representations being used for ligand-based virtual screening, topological fingerprints encoding the presence of substructural fragments in molecules have been by far the most commonly used (Hert *et al.*, 2004). Still at the topological level, the use of distributions of atom-centred feature pairs has also been proven to be highly effective in a variety of virtual screening applications (Schneider *et al.*, 1999; Schuffenhauer *et al.*, 2003; Gregori-Puigjane and Mestres, 2006). In contrast to topological approaches, methods based on geometrical representations of molecular structures can be used instead. Among them, flexible superposition of molecules onto one or multiple conformations of a reference bioactive ligand is a well-established methodology in virtual screening (Lemmen and Lengauer, 2000; Mestres and Veeneman, 2003; Jain, 2004).

Perhaps the most widely employed methods requiring 3D structure representations of molecules are those exploiting the concept of pharmacophore similarity (Mason *et al*., 2001). By definition, a pharmacophore is the 3D arrangement of molecular features necessary for bioactivity (Wermuth *et al*., 1998) and the underlying methodologies have been widely described (Martin, 1992, 1993; Guner, 2000; Langer and Hoffman, 2006). Although initially slow to gain an industrial foothold, pharmacophore approaches have subsequently been applied to many therapeutic targets for the virtual screening of compound databases (Sprague, 1995; Barnum *et al*., 1996; Sprague and Hoffman, 1997). Successful applications of the use of pharmacophores in virtual screening include the identification of hits for a variety of targets such as protein kinase C (Wang *et al*., 1994), farnesyltransferase (Kaminski *et al*., 1997), HIV integrase (Nicklaus *et al*., 1997; Carlson *et al*., 2000), endothelial differentiation gene receptor antagonists (Koide *et al*., 2002), urotensin antagonists (Flohr *et al*., 2002), CCR5 antagonist (Debnath, 2003) and mesangial cell proliferation inhibitor discovery (Kurogi *et al*., 2001), to mention a few. Pharmacophores have also been generated for numerous ADME/Tox-related proteins (Ekins and Swaan, 2004). These efforts suggest that pharmacophore-based approaches may have considerable versatility and applicability to be used with difficult biological targets. Newer methods for extracting ligand pharmacophores from protein cavities have also emerged recently (Wolber and Langer, 2005) which may facilitate the generation of pharmacophores for multiple targets and simultaneous pharmacophore selectivity screening.

Target-based methods. Target-based virtual screening methods depend on the availability of structural information of the target, that being either determined experimentally or derived computationally by means of homology modelling techniques (Shoichet, 2004; Klebe, 2006). These methods aim at providing, on one hand, a good approximation of the expected conformation and orientation of the ligand into the protein cavity (docking) and, on the other hand, a reasonable estimation of its binding affinity (scoring). Despite its appealing concept, docking and scoring ligands in target sites is still a challenging process after more than 20 years of research in the field (Kitchen *et al*., 2004; Ghosh *et al*., 2006; Leach *et al*., 2006) and the performance of different implementations has been found to vary widely depending on the given target (Cummings *et al*., 2005). To alleviate this situation, the use of multiple active site corrections has been suggested to remedy the ligand-dependent biases in scoring functions (Vigers and Rizzi, 2004) and the use of multiple scoring functions (consensus scoring) has been also recommended to improve the enrichment of true positives in virtual screening (Charifson *et al*., 1999). Also, as the number of protein–ligand complexes available continues to grow, docking methods are beginning to incorporate all the information derived from the conformation adopted by protein-bound ligands as a knowledge-based strategy to correct some of the limitations of current scoring functions and actively guide the orientation of the ligands into the protein cavity (Fradera and Mestres, 2004).

In spite of all these limitations, target-based virtual screening has gained a reputation in successfully identifying and generating novel bioactive compounds. As an example, the use of a knowledge-based potential (SMoG) in protein–ligand docking, resulted in the identification of new picomolar ligands for the human carbonic anhydrase II (Grzybowski *et al*., 2002). Docking methods have also resulted in the discovery of novel inhibitors for several kinase targets, including cyclin-dependent kinases, epidermal growth factor receptor kinase and vascular endothelial growth factor receptor 2 kinase among others (Muegge and Enyedy, 2004). Finally, the application of docking methods to targets for which experimentally determined structures are not available yet has gained considerable attention in recent years, particularly for the many targets of therapeutic relevance belonging to the superfamily of G-protein-coupled receptors (GPCRs) (Bissantz *et al*., 2003). In these cases, structural information is generated computationally by modelling the structure of the target of interest on the basis of a template structure of a related target, including often information on ligands as restraints (Evers *et al*., 2003). Such strategies have resulted in the successful identification of novel antagonists for the neurokinin-1 and the α1A-adrenergic receptors (Evers and Klebe, 2004; Evers and Klabunde, 2005b).

Virtual affinity profiling

If virtual ligand screening extended QSAR along the chemical dimension, recent trends in virtual affinity profiling are adding a further biological dimension to it. A wave of new methods that are capable of estimating the pharmacological profile of molecules on multiple targets have been recently reported. These promise to have a strong influence in drug discovery as a means for detecting during the optimisation process potential side effects of compounds due to off-target affinities (O'Connor and Roth, 2005; Paolini *et al*., 2006). This notwithstanding, it should be recognised that the current flourishing of these methods is mainly a consequence of the important progress experienced by some coordinated initiatives dedicated to data collection, classification and storage, both in extracting pharmacological data for ligands as well as in gathering structural information for proteins (Mestres, 2004).

Ligand-based methods. The development of ligand-based affinity profiling methods has benefited enormously from the construction of annotated chemical libraries that incorporate literature-based pharmacological data into traditional chemical repositories (Savchuk *et al*., 2004). Among these, the WOMBAT database (Sunset Molecular Discovery LLC, Santa Fe, NM, USA) provides biological information for 120 400 molecules reported in medicinal chemistry journals over the last 30 years. The MDL Drug Data Report or MDDR (MDL Information Systems, San Ramon, CA, USA) includes information on therapeutic action and biological activity for over 132 000 compounds gathered from patent literature, journals and congresses. The AurSCOPE databases (Aureus Pharma, Paris, France) offer a collection of chemical libraries containing over 320 000 molecules annotated to about

1 300 000 biological activities related to members of therapeutically relevant protein families covered in more than 38 000 publications. And finally the MedChem and Target Inhibitor databases (GVK Biosciences, Hyderabad, India) compile around 2 000 000 molecules with biological activity, toxicity and pharmacological information for therapeutically relevant protein families extracted from more than 20 000 publications. All these massive annotation initiatives ultimately allow the end-user to connect small molecules to target proteins on the basis of data published in scientific publications. This is then available to use in creating ligand-based protein models that can be used for virtual affinity profiling (Schuffenhauer and Jacoby, 2004).

One of the earliest developed initiatives is the computer system PASS (Poroikov *et al*., 2000), which is based on the analysis of structure–activity relationships for a training set of compounds consisting of about 35 000 biologically active compounds extracted from the literature. The system provides a prediction of the activity spectra of substances for more than 500 biological activities. In a second similar pioneering work, the relationships between the chemical structures of 48 compounds and their pharmacological profile against a set of more than 70 receptors, transporters and channels relevant to a central nervous system (CNS)-oriented project were analysed (Poulain *et al*., 2001). Along the same line, a biospectra similarity analysis was performed by clustering a set of 1567 drugs for which percent inhibition values determined at single high ligand concentration was available for a set of 92 assays (Fliri *et al*., 2005). In another study, the MDDR database was used as a source of ligands annotated to the four major target classes, namely, enzymes, GPCRs, nuclear receptors and ligand-gated ion channels (Schuffenhauer *et al*., 2002). The resulting ligand-target classification scheme was subsequently used for searching for structures binding to dopamine D2, all dopamine receptors, and all amine-binding class A GPCRs using dopamine D2-binding compounds as a reference set. The WOMBAT database has also been used to derive a multiple-category Laplacian-modified naïve Bayesian model trained on extended connectivity fingerprints for a set of 964 target classes (Nidhi *et al*., 2006). The model was then applied to predict the top three most likely protein targets for compounds from the MDDR database and it was found that on average, it was 77% correct at target identification. A database containing thousands of substructures annotated to over 500 biological endpoints, including pharmacological, cell-based, animal model-based, toxicity, ADME and therapeutic outcomes, has been reported (Merlot *et al*., 2003). The substructural database is then used as an end point alert for molecules containing any of the substructures catalogued in the database associated with a given end point. From a similar perspective, a database of topological scaffolds was recently constructed (Wilkens *et al*., 2005) in which each scaffold is linked to the distribution of affinities associated to all molecules containing it. This can be used as a means to identify those scaffolds showing a statistical enrichment in certain biological activities. In a further example, the scaffolds extracted from a literature-based chemical library of 1426 molecules annotated to 27 members of the nuclear receptor family were used to identify the set of most promiscuous scaffolds within this family (Cases *et al*., 2005). Similarly, an ontology-based pattern identification algorithm was used to identify nearly 1500 scaffold families with statistically significant structure–activity profile relationships (Yan *et al*., 2006). The use of multiple structure-based pharmacophore models built for various viral targets has also been investigated as a ligand-based affinity profiling method to estimate the pharmacological profile of 100 antiviral compounds (Steindl *et al*., 2006). The results showed successful virtual activity profiling for approximately 90% of all input molecules. A final example used a novel set of descriptors based on topological atom-centred feature-based distributions (Gregori-Puigjane and Mestres, 2006) and a database of 2033 molecules annotated to 25 nuclear receptors to derive a ligand-based descriptor model for the profiling of chemical libraries on the family of nuclear receptors (Mestres *et al*., 2006). Application of this ligand-based nuclear receptor profiling approach to an additional database of 2944 drugs provided suggestions of potential off-targets affinities.

Target-based methods. The development of target-based affinity profiling methods has taken advantage of the functional coverage of protein families provided by the almost exponential growth in the number of experimentally determined protein structures (Mestres, 2005). Unfortunately, primarily because of technical difficulties (for example, membrane-bound proteins), not all of the therapeutically relevant protein families are at present equally covered by 3D structures. For example, with over 20 000 entries, enzymes are by far the structurally most populated family (Garcia-Serna *et al*., 2006). In contrast, around 200 structures are available for nuclear receptors and ligand-gated ion channels, whereas only a handful has been resolved for GPCRs. In the latter case, homology modelling techniques are required to complement current low coverage levels of experimentally determined structures with computationally derived structural models (Pieper *et al*., 2006).

Although extremely computationally demanding compared with ligand-based methods, applications of target-based virtual profiling (also referred to as inverse docking) have been reported in recent years (Toledo-Sherman and Chen, 2002). For example, reasonable predictions for the scaffold and S1-specificity preferences for serine proteases were obtained when multiple combinatorial libraries were docked against trypsin, chymotrypsin and elastase (Lamb *et al*., 2001). The same approach was further applied to virtually screen three libraries against a panel of six purine phosphoribosyltransferases (PRTs) from different species, giving rise to the discovery of micromolar inhibitors of *Giardia lamblia* guanine PRT (GPRT) that displayed up to sevenfold selectivity when tested in human GPRT (Aronov *et al*., 2001). Systematic virtual screening of a library consisting of 5000 random compounds and 78 known active ligands against 19 different protein structures representative of 10 nuclear receptors was also reported (Schapira *et al*., 2003). Enrichments of between 33- to 100-fold were obtained for all but one receptor and revealed that, for a particular ligand, it is possible to identify the correct target within the receptor family. However, reliable discrimination

between the closely related receptor isoforms remained a challenge. Application of an inverse docking approach involving multiple-conformer shape-matching alignment of a molecule to a protein cavity on two therapeutic agents showed that 50% of the computer-identified potential protein targets were implicated or confirmed by experiments (Chen and Zhi, 2001b). The same approach was also used to forecast potential toxicity problems and for the identification of protein targets implicated in side effects of small molecules, with an estimated prediction success rate of 83% (Chen and Ung, 2001a). The sc-PDB is a collection of 6415 3D structures of binding sites extracted from the Protein Data Bank (Kellenberger *et al*., 2006). Exploiting this binding site database, the native target of four unrelated ligands could be identified among the top 1% of scored binding sites, with 70 to 100-fold enrichment relative to random screening (Paul *et al*., 2004). A method has also been reported for rapidly computing the relative affinity of inhibitors to individual members of the kinase family (Rockey and Elcock, 2005). This was tested on five known kinase inhibitors, and was able to identify the correct native targets of inhibitors as well as reproducing the experimental trends in binding affinities. A web server for inverse docking was recently reported that allows automatic screening of small molecules over a target database of 698 protein structures covering 15 therapeutic areas (Li *et al*., 2006).

Data visualisation

Computational methods have the potential to generate predictions for many different types of pharmacological and physicochemical properties for each molecule structure, the analysis of such data would indicate the need for multidimensional methods and perhaps sophisticated visualisation tools for data mining (Cheng *et al*., 2002; Ekins *et al*., 2002, 2006). Commercially available tools such as Diva and Spotfire (Ahlberg, 1999) have been widely used for analysis of ADME and physicochemical property data (Ekins *et al*., 2002, 2006; Stoner *et al*., 2004) or integrated into proprietary decision support systems (Rojnuckarin *et al*., 2005), whereas newer methods are also available (Oellien *et al*., 2005) with similar 3D graphing and filtering options. Further methods such as agglomerative hierarchical clustering based on 2D structural similarity, recursive partitioning, Sammon maps, self-organising maps and generative topographic mapping could be used with computational predictions (Balakin *et al*., 2005; Kibbey and Calvet, 2005; Maniyar *et al*., 2006; Yamashita *et al*., 2006).

Because molecules may have multiple off-target effects simultaneously it will be important to understand how they perturb the proteome either alone or in a combination (Sharom *et al*., 2004). One way to visualise the target–molecule interactions would be as a network of proteins and small molecules represented as nodes connected by edges when an interaction above a particular affinity exists between them (Ekins *et al*., 2005a, b) or alternatively a target-based network in which the proteins are highlighted when they are shown to interact with a small molecule (Ekins, 2006). Such network analysis has traditionally been used for putting the genes with expression data into the context of their known pathways of transcription factors, serving as a method to understand some of the complex interactions. For example, commercially available tools can be used to mine biological knowledge to create networks for molecules alone which may be useful to understand the relationships between endogenous molecules (Figure 3a) or they can facilitate visualising gene expression data for

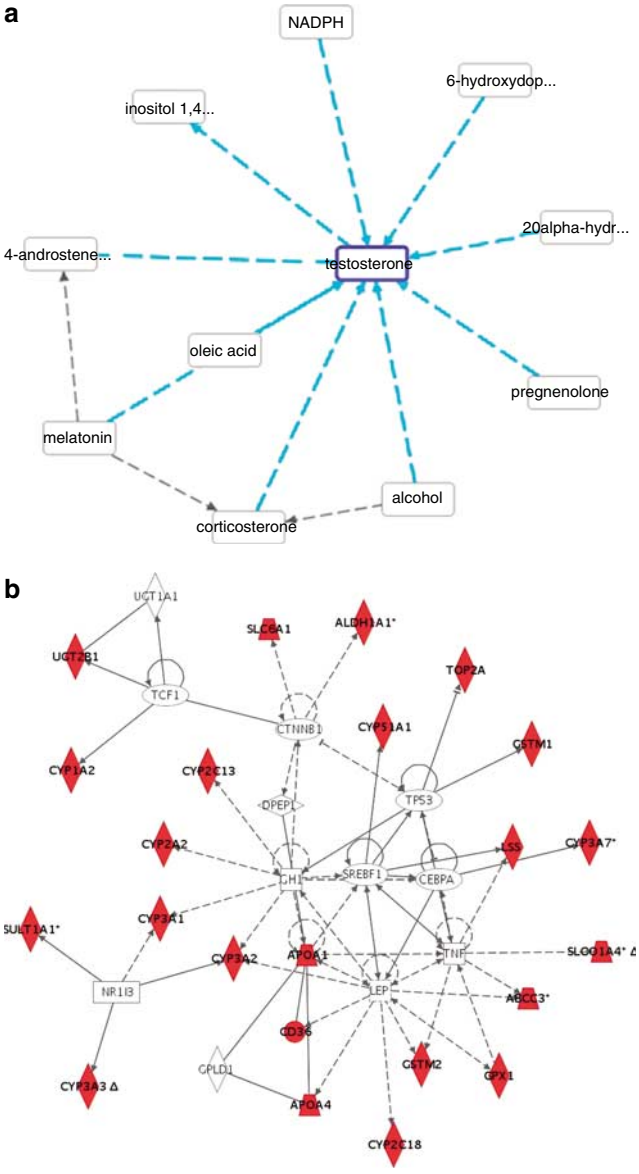


Figure 3 (**a**) An endogenous molecule network generated using Ingenuity Pathways Analysis (Ingenuity Systems Inc., Redwood City, CA, USA). Solid lines represent direct interactions and dashed lines represent indirect interactions. (**b**) A network showing the connectivity via direct interactions of several key nuclear hormone receptors (rectangles) and their regulation of several transporters (trapezoid) and enzymes (diamonds) involved in drug absorption and metabolism while gene expression data from rats after treatment with 2(S)-((3,5-bis(trifluoromethyl)benzyl)-oxy)-3(S) phenyl-4-((3-oxo-1,2,4-tri azol-5-yl)methyl)morpholine (L-742694) is overlaid (Hartley *et al*., 2004). The network shows key upregulated transporter and enzyme genes (red symbols). Note that several genes are connected to PXR (NRI13).

nuclear receptors that regulate drug metabolism and toxicity (Figure 3b).

This type of network approach may also help in designing drugs with affinity for multiple targets (Csermely *et al*., 2005) or avoid anti-targets. For example, an interaction network between 25 nuclear receptors was recently constructed on the basis of an annotated chemical library containing 2033 molecules (Mestres *et al*., 2006). The network revealed potential cross-pharmacologies between members of this family with implications for the side-effect prediction of small molecules. We have discussed previously some of the challenges ahead for such network approaches including comparing multiple networks (Ekins *et al*., 2006). This could be important if we are to use this type of approach for visualising the effect of a molecule on the connected proteome and the comparison of related molecule effects or different doses. More complex simulations of network biology may also be important to optimise targeting, dosing level and frequency. An *in silico* simulation of inhibitor effects (incorporating PK data) on the NF-κB pathway has shown a greater potential for protein oscillatory behaviour for inhibition upstream of this protein than for direct inhibition. Such behaviour is common in networks with negative feedback and should be considered and understood for maximal therapeutic benefit (Sung and Simon, 2004).

Summary

The first part of this review has briefly described the development of *in silico* pharmacology through the development of methods including databases, quantitative structure–activity relationships, similarity searching, pharmacophores, homology models and other molecular modelling, machine learning, data mining, network analysis tools and data analysis tools that use a computer. We have introduced how some of these methods can be used for virtual ligand screening and virtual affinity profiling. Although these methods are not proven yet to 'discover drugs' alone, they represent progress by increasingly demonstrating their ability to deliver enrichment in identifying active molecules for the target of interest when compared with random selection or other traditional methods. In the accompanying second part of the review, we shall describe in more detail the successful ligand screening efforts for specific target classes and we will discuss some of the advantages and disadvantages of *in silico* methods with respect to *in vitro* and *in vivo* methods for pharmacology research.

Acknowledgements

SE gratefully acknowledges Dr Cheng Chang and Dr Peter W Swaan (University of Maryland), Dr Konstantin V Balakin (Chemical Diversity, Inc.) for *in silico* pharmacology collaborations over the past several years and Dr Hugo Kubinyi for his insightful efforts in tabulating the successful applications of *in silico* approaches which was inspirational. Ingenuity Systems Inc., graciously provided access to Ingenuity Pathways Analysis. SE kindly acknowledges Dr Maggie AZ Hupcey for her support. JM acknowledges the research funding provided by the Spanish Ministerio de Educación y Ciencia (project reference BIO2005-04171) and the Instituto de Salud Carlos III.

Conflict of interest

The authors state no conflict of interest.

References

Ahlberg C (1999). Visual exploration of HTS databases: bridging the gap between chemistry and biology. *Drug Discov Today* **4**: 370–376.
Akamatsu M (2002). Current state and perspectives of 3D-QSAR. *Curr Top Med Chem* **2**: 1381–1394.
Albert A (1971). Relations between molecular structure and biological activity: stages in the evolution of current concepts. *Ann Rev Pharmacol* **11**: 13–36.
Albert A (1985). *Selective Toxicity. The Physcico-Chemical Basis of Therapy*. Chapman and Hall: London.
Aradi I, Erdi P (2006). Computational neuropharmacology: dynamical approaches in drug discovery. *Trends Pharmacol Sci* **27**: 240–243.
Arïens EJ (1979). Receptors: from fiction to fact. *Trends Pharmacol Sci* **1**: 11–15.
Aronov AM, Munagala NR, Kuntz ID, Wang CC (2001). Virtual screening of combinatorial libraries across a gene family: in search of inhibitors of *Giardia lamblia* guanine phosphoribosyltransferase. *Antimicrob Agents Chemother* **45**: 2571–2576.
Bajorath J (2002). Integration of virtual and high-throughput screening. *Nat Rev Drug Disc* **1**: 882–894.
Balakin KV, Ivanenkov YA, Savchuk NP, Ivaschenko AA, Ekins S (2005). Comprehensive computational assessment of ADME properties using mapping techniques. *Curr Drug Disc Tech* **2**: 99–113.
Barnum D, Greene J, Smellie A, Sprague P (1996). Identification of common functional configurations among molecules. *J Chem Inf Comput Sci* **36**: 563–571.
Bissantz C, Bernard P, Hibert M, Rognan D (2003). Protein-based virtual screening of chemical databases. II. Are homology models of G-protein coupled receptors suitable targets? *Proteins* **50**: 5–25.
Bleicher KH, Bohm HJ, Muller K, Alanine AI (2003). Hit and lead generation: beyond high-throughput screening. *Nat Rev Drug Discov* **2**: 369–378.
Burgen ASV (1981). Conformational changes and drug action. *Fed Proc* **40**: 2723–2728.
Button WG, Judson PN, Long A, Vessey JD (2003). Using absolute and relative reasoning in the prediction of the potential metabolism of xenobiotics. *J Chem Inf Compu Sci* **43**: 1371–1377.
Carlson HA, Masukawa KM, Rubins K, Bushman FD, Jorgenson WL, Lins RD *et al*. (2000). Developing a dynamic pharmacophore model for HIV-1 integrase. *J Med Chem* **43**: 2100–2114.
Cases M, Garcia-Serna R, Hettne K, Weeber M, van der Lei J, Boyer S *et al*. (2005). Chemical and biological profiling of an annotated compound library directed to the nuclear receptor family. *Curr Top Med Chem* **5**: 763–772.
Charifson PS, Corkery JJ, Murcko MA, Walters WP (1999). Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem* **42**: 5100–5109.
Chen YZ, Ung CY (2001a). Prediction of potential toxicity and side effect protein targets of a small molecule by a ligand–protein inverse docking approach. *J Mol Graph Model* **20**: 199–218.
Chen YZ, Zhi DG (2001b). Ligand–protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins* **43**: 217–226.
Cheng A, Diller DJ, Dixon SL, Egan WJ, Lauri G, Merz Jr KMJ (2002). Computation of the physico-chemical properties and data mining of large molecular collections. *J Comput Chem* **23**: 172–183.

Chohan KK, Paine SW, Waters NJ (2006). Quantitative structure activity relationships in drug metabolism. *Curr Top Med Chem* **6**: 1569–1578.

Cramer RD, Patterson DE, Bunce JD (1988). Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc* **110**: 5959–5967.

Csermely P, Agoston V, Pongor S (2005). The efficiency of multi-target drugs: the network approach might help drug design. *Trends Pharmacol Sci* **26**: 178–182.

Cummings MD, DesJarlais RL, Gibbs AC, Mohan V, Jaeger EP (2005). Comparison of automated docking programs as virtual screening tools. *J Med Chem* **48**: 962–976.

Cushny A (1926). *Biological Relations of Optical Isomeric Substances*. Williams and Wilkins: Baltimore.

Danchin A (2002). *The Delphic Boat – What Genomes Tell Us*. Harvard University Press: Cambridge, MA.

Danchin A, Medigue C, Gascuel O, Soldano H, Henaut A (1991). From data banks to data bases. *Res Microbiol* **142**: 913–916.

Darvas F, Marokhazi S, Kormos P, Kulkarni P, Kalasz G, Papp A (1999). MetabolExpert: its use in metabolism research and in combinatorial chemistry. In: Erhardt PW (ed). *Drug Metabolism: Databases and High-Throughput Testing During Drug Design and Development*. International Union of Pure and Applied Chemistry and Blackwell Science: London.

Debnath AK (2003). Generation of predictive pharmacophore models for CCR5 anatagonists: study with piperidine- and piperazine-based compounds as a new class of HIV-1 entry inhibitors. *J Med Chem* **46**: 4501–4515.

Dimitrov S, Dimitrova G, Pavlov T, Dimitrova N, Patlewicz G, Niemela J *et al*. (2005). A stepwise approach for defining the applicability domain of SAR and QSAR models. *J Chem Inf Model* **45**: 839–849.

Dudek AZ, Arodz T, Galvez J (2006). Computational methods in developing quantitative structure–activity relationships (QSAR): a review. *Comb Chem High Throughput Screen* **9**: 213–228.

Ekins S (2006). Systems-ADME/Tox: resources and network approaches. *J Pharmacol Toxicol Methods* **53**: 38–66.

Ekins S, Boulanger B, Swaan PW, Hupcey MAZ (2002). Towards a new age of virtual ADME/TOX and multidimensional drug discovery. *J Comput Aided Mol Des* **16**: 381–401.

Ekins S, Kirillov E, Rakhmatulin EA, Nikolskaya T (2005a). A novel method for visualizing nuclear hormone receptor networks relevant to drug metabolism. *Drug Metab Dispos* **33**: 474–481.

Ekins S, Mestres J, Testa B (2007). *In silico* pharmacology for drug discovery: applications to targets and beyond. *Br J Pharmacol* [E-pub ahead of print: 4 June 2007; doi:10.1038/sj.bjp.0707306].

Ekins S, Nikolsky Y, Nikolskaya T (2005b). Techniques: application of systems biology to absorption, distribution, metabolism, excretion, and toxicity. *Trends Pharmacol Sci* **26**: 202–209.

Ekins S, Shimada J, Chang C (2006). Application of data mining approaches to drug delivery. *Adv Drug Del Rev* **58**: 1409–1430.

Ekins S, Swaan PW (2004). Development of computational models for enzymes, transporters, channels and receptors relevant to ADME/TOX. *Rev Comp Chem* **20**: 333–415.

Evers A, Gohlke H, Klebe G (2003). Ligand-supported homology modelling of protein binding-sites using knowledge-based potentials. *J Mol Biol* **334**: 327–345.

Evers A, Hessler G, Matter H, Klabunde T (2005a). Virtual screening of biogenic amine-binding G-protein coupled receptors: comparative evaluation of protein- and ligand-based virtual screening protocols. *J Med Chem* **48**: 5448–5465.

Evers A, Klabunde T (2005b). Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the alpha1A adrenergic receptor. *J Med Chem* **48**: 1088–1097.

Evers A, Klebe G (2004). Successful virtual screening for a sub-micromolar antagonist of the neurokinin-1 receptor based on a ligand-supported homology model. *J Med Chem* **47**: 5381–5392.

Fliri AF, Loging WT, Thadeio PF, Volkmann RA (2005). Biospectra analysis: model proteome characterizations for linking molecular structure and biological response. *J Med Chem* **48**: 6918–6925.

Flohr S, Kurz M, Kostenis E, Brkovich A, Fournier A, Klabunde T (2002). Identification of nonpeptidic urotensin II receptor antagonists by virtual screening based on a pharmacophore model derived from structure–activity relationships and nuclear magnetic resonance studies on urotensin II. *J Med Chem* **45**: 1799–1805.

Fradera X, Mestres J (2004). Guided docking approaches to structure-based design and screening. *Curr Top Med Chem* **4**: 687–700.

Garcia-Serna R, Opatowski L, Mestres J (2006). FCP: functional coverage of the proteome by structures. *Bioinformatics* **22**: 1792–1793.

Ghosh S, Nie A, An J, Huang Z (2006). Structure-based virtual screening of chemical libraries for drug discovery. *Curr Opin Chem Biol* **10**: 194–202.

Gohlke H, Klebe G (2002). Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew Chem Int Ed Engl* **41**: 2644–2676.

Gozalbes R, Doucet JP, Derouin F (2002). Application of topological descriptors in QSAR and drug design: history and new trends. *Curr Drug Targets Infect Disord* **2**: 93–102.

Gregori-Puigjane E, Mestres J (2006). SHED: Shannon entropy descriptors from topological feature distributions. *J Chem Inf Model* **46**: 1615–1622.

Grzybowski BA, Ishchenko AV, Kim C-K, Topalov G, Chapman R, Christianson DW *et al*. (2002). Combinatorial computational method gives new picomolar ligands for a known enzyme. *Proc Natl Acad Sci USA* **99**: 1270–1273.

Guner OF (2000). *Pharmacophore, Perception, Development, and Use in Drug Design*. University International Line: San Diego.

Hansch C (1972). Quantitative relationships between lipophilic character and drug metabolism. *Drug Metab Rev* **1**: 1–14.

Hansch C, Fujita T (1964). Rho-sigma-pi analysis. A method for the correlation of biological activity and chemical structure. *J Am Chem Soc* **86**: 1616–1626.

Hansch C, Hoekman D, Leo A, Weininger D, Selassie CD (2002). Chem-bioinformatics: comparative QSAR at the interface between chemistry and biology. *Chem Rev* **102**: 783–812.

Hartley DP, Dai X, He YD, Carlini EJ, Wang B, Huskey SE *et al*. (2004). Activators of the rat pregnane X receptor differentially modulate hepatic and intestinal gene expression. *Mol Pharmacol* **65**: 1159–1171.

Hert J, Willett P, Wilton DJ, Acklin P, Azzaoui K, Jacoby E *et al*. (2004). Comparison of topological descriptors for similarity-based virtual screening using multiple bioactive reference structures. *Org Biomol Chem* **2**: 3256–3266.

Jain AN (2004). Ligand-based structural hypotheses for virtual screening. *J Med Chem* **47**: 947–961.

Johnson M, Maggiora GM (2006). *Concepts and Applications of Molecular Similarity*. John Wiley & Sons: New York.

Kaminski JJ, Rane DF, Snow ME, Weber L, Rothofsky ML, Anderson SDSLL (1997). Identification of novel farnesyl protein transferase inhibitors using three-dimensional searching methods. *J Med Chem* **40**: 4103–4112.

Karelson M (2000). *Molecular Descriptors in QSAR/QSPR*. Wiley-VCH: New York.

Kellenberger E, Muller P, Schalon C, Bret G, Foata N, Rognan D (2006). sc-PDB: an annotated database of druggable binding sites from the Protein Data Bank. *J Chem Inf Model* **46**: 717–727.

Kibbey C, Calvet A (2005). Molecular Property eXplorer: a novel approach to visualizing SAR using tree-maps and heatmaps. *J Chem Inf Model* **45**: 523–532.

Kitchen DB, Decornez H, Furr JR, Bajorath J (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* **3**: 935–949.

Klebe G (1998). Comparative molecular similarity indices analysis: CoMSIA. *Perspect Drug Disc Des* **12–14**: 87–104.

Klebe G (2006). Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov Today* **11**: 580–594.

Klopman G, Tu M (1999). META: a program for the prediction of the products of mammal metabolism of xenobiotics. In: Erhardt PW (ed). *Drug Metabolism: Databases and High-Throughput Testing During Drug Design and Development*. International Union of Pure and Applied Chemistry and Blackwell Science: London.

Koide Y, Hasegawa T, Takahashi A, Endo A, Mochizuki N, Nakagawa M *et al*. (2002). Development of novel EDG3 antagonists using a 3D database search and their structure–activity relationships. *J Med Chem* **45**: 4629–4638.

Kubinyi H, Folkers G, Martin YC (1998). *3D-QSAR in Drug Design*. Kluwer/ESCOM: Leiden.
Kulkarni SA, Zhu J, Blechinger S (2005). *In silico* techniques for the study and prediction of xenobiotic metabolism: a review. *Xenobiotica* **35**: 955–973.
Kurogi Y, Miyata K, Okamura T, Hashimoto K, Tsutsumi K, Nasu M *et al*. (2001). Discovery of novel mesangial cell proliferation inhibitors using a three-dimensional database searching method. *J Med Chem* **44**: 2304–2307.
Kurup A (2003). C-QSAR: a database of 18,000 QSARs and associated biological and physical data. *J Comput Aided Mol Des* **17**: 187–196.
Lahana R (1999). How many leads from HTS? *Drug Discov Today* **4**: 447–448.
Lamb ML, Burdick KW, Toba S, Young MM, Skillman AG, Zou X *et al*. (2001). Design, docking, and evaluation of multiple libraries against multiple targets. *Proteins* **42**: 296–318.
Langer T, Hoffman RD (2006). *Pharmacophores and Pharmacophore Searches*. Wiley-VCH: Weinheim.
Langowski J, Long A (2002). Computer systems for the prediction of xenobiotic metabolism. *Adv Drug Del Rev* **54**: 407–415.
Leach AR, Shoichet BK, Peishoff CE (2006). Prediction of protein–ligand interactions. Docking and scoring: successes and gaps. *J Med Chem* **49**: 5851–5855.
Lemmen C, Lengauer T (2000). Computational methods for the structural alignment of molecules. *J Comput Aided Mol Des* **14**: 215–232.
Lengauer T, Lemmen C, Rarey M, Zimmermann M (2004). Novel technologies for virtual screening. *Drug Discov Today* **9**: 27–34.
Li H, Gao Z, Kang L, Zhang H, Yang K, Yu K *et al*. (2006). TarFisDock: a web server for identifying drug targets with docking approach. *Nucleic Acids Res* **34**: W219–W224.
Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Del Rev* **23**: 3–25.
Maniyar DM, Nabney IT, Williams BS, Sewing A (2006). Data visualization during the early stages of drug discovery. *J Chem Inf Model* **46**: 1806–1818.
Mannens GSJ, Snel CAW, Hendrickx J, Verhaeghe T, Le Jeune L, Bode W *et al*. (2002). The metabolism and excretion of galantamine in rats, dogs, and humans. *Drug Metab Dispos* **30**: 553–563.
Martin YC (1992). 3D database searching in drug design. *J Med Chem* **35**: 2145–2154.
Martin YC, Bures MG, Danaher EA, DeLazzer J, Lico I, Pavlik PA (1993). A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists. *J Comput Aided Mol Des* **7**: 83–102.
Mason JS, Good AC, Martin EJ (2001). 3D pharmacophores in drug discovery. *Curr Pharm Des* **7**: 567–597.
Merlot C, Domine D, Cleva C, Church DJ (2003). Chemical substructures in drug discovery. *Drug Discov Today* **8**: 594–602.
Mestres J (2004). Computational chemogenomics approaches to systematic knowledge-based drug discovery. *Curr Opin Drug Discov Dev* **7**: 304–313.
Mestres J (2005). Representativity of target families in the Protein Data Bank: impact for family-directed structure-based drug discovery. *Drug Discov Today* **10**: 1629–1637.
Mestres J, Couce-Martin L, Gregori-Puigjane E, Cases M, Boyer S (2006). Ligand-based approach to in silico pharmacology: nuclear receptor profiling. *J Chem Inf Model* **46**: 2725–2736.
Mestres J, Veeneman GH (2003). Identification of 'latent hits' in compound screening collections. *J Med Chem* **46**: 3441–3444.
Meyer H (1899). Zur Theorie der Alkoholnarkose. *Arch Expl Pathol Pharmakol* **42**: 110–118.
Muegge I, Enyedy IJ (2004). Virtual screening for kinase targets. *Curr Med Chem* **11**: 693–707.
Nicklaus MC, Neamati N, Hong H, Mazumder A, Sunder S, Chen J *et al*. (1997). HIV-1 integrase pharmacophore: discovery of inhibitors through three-dimensional database searching. *J Med Chem* **40**: 920–929.
Nidhi, Glick M, Davies JW, Jenkins JL (2006). Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. *J Chem Inf Model* **46**: 1124–1133.
O'Connor KA, Roth BL (2005). Finding new tricks for old drugs: an efficient route for public-sector drug discovery. *Nat Rev Drug Discov* **4**: 1005–1014.
Oellien F, Ihlenfeldt WD, Gasteiger J (2005). InfVis – platform-independent visual data mining of multidimensional chemical data sets. *J Chem Inf Model* **45**: 1456–1467.
Oprea TI, Matter H (2004). Integrating virtual screening in lead discovery. *Curr Opin Chem Biol* **8**: 349–358.
Ortiz AR, Pisabarro MT, Gago F, Wade RC (1995). Prediction of drug binding affinities by comparative binding energy analysis. *J Med Chem* **38**: 2681–2691.
Overton E (1901). *Studien über die Narkose*. Gustav Fischer: Jena.
Paolini GV, Shapland RH, van Hoorn WP, Mason JS, Hopkins AL (2006). Global mapping of pharmacological space. *Nat Biotechnol* **24**: 805–815.
Parascandola J (1980). Origins of the receptor theory. *Trends Pharmacol Sci* **1**: 189–192.
Pastor M, Cruciani G, McLay I, Pickett S, Clementi S (2000). GRid-INdependent descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors. *J Med Chem* **43**: 3233–3243.
Paul N, Kellenberger E, Bret G, Muller P, Rognan D (2004). Recovering the true targets of specific ligands by virtual screening of the protein data bank. *Proteins* **54**: 671–680.
Perkins R, Fang H, Tong W, Welsh WJ (2003). Quantitative structure–activity relationship methods: perspectives on drug discovery and toxicology. *Environ Toxicol Chem* **22**: 1666–1679.
Pieper U, Eswar N, Davis FP, Braberg H, Madhusudhan MS, Rossi A *et al*. (2006). MODBASE: a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res* **34**: D291–D295.
Polanski J, Bak A, Gieleciak R, Magdziarz T (2006). Modeling robust QSAR. *J Chem Inf Model* **46**: 2310–2318.
Poroikov VV, Filimonov DA, Borodina YV, Lagunin AA, Kos A (2000). Robustness of biological activity spectra predicting by computer program PASS for noncongeneric sets of chemical compounds. *J Chem Inf Comput Sci* **40**: 1349–1355.
Poulain R, Horvath D, Bonnet B, Eckhoff C, Chapelain B, Bodinier MC *et al*. (2001). From hit to lead. Analyzing structure–profile relationships. *J Med Chem* **44**: 3391–3401.
Rockey WM, Elcock AH (2005). Rapid computational identification of the targets of protein kinase inhibitors. *J Med Chem* **48**: 4138–4152.
Rojnuckarin A, Gschwend DA, Rotstein SH, Hartsough DS (2005). ArQiologist: an integrated decision support tool for lead optimization. *J Chem Inf Model* **45**: 2–9.
Savchuk NP, Balakin KV, Tkachenko SE (2004). Exploring the chemogenomic knowledge space with annotated chemical libraries. *Curr Opin Chem Biol* **8**: 412–417.
Schapira M, Abagyan R, Totrov M (2003). Nuclear hormone receptor targeted virtual screening. *J Med Chem* **46**: 3045–3059.
Schneider G, Neidhart W, Giller T, Schmid G (1999). 'Scaffold-Hopping' by topological pharmacophore search: a contribution to virtual screening. *Angew Chem Int Ed Engl* **38**: 2894–2896.
Schuffenhauer A, Floersheim P, Acklin P, Jacoby E (2003). Similarity metrics for ligands reflecting the similarity of the target proteins. *J Chem Inf Comput Sci* **43**: 391–405.
Schuffenhauer A, Jacoby E (2004). Annotating and mining the ligand-target chemogenomics knowledge space. *Drug Disc Today Biosilico* **2**: 190–200.
Schuffenhauer A, Zimmermann J, Stoop R, van der Vyver JJ, Lecchini S, Jacoby E (2002). An ontology for pharmaceutical ligands and its application for *in silico* screening and library design. *J Chem Inf Comput Sci* **42**: 947–955.
Sharom JR, Bellows DS, Tyers M (2004). From large networks to small molecules. *Curr Opin Chem Biol* **8**: 81–90.
Shimada J (2006). The challenges of making useful protein-ligand free energy predictions for drug discovery. In: Ekins S (ed). *Computer Applications in Pharmaceutical Research and Development*. John Wiley and Sons: Hoboken, pp 321–351.
Shoichet BK (2004). Virtual screening of chemical libraries. *Nature* **432**: 862–865.
Sieburg HB (1990). Physiological studies *in silico*. *Studies in the Sciences of Complexity* **12**: 321–342.

Sippl W (2002). Development of biologically active compounds by combining 3D QSAR and structure-based design methods. *J Comput Aided Mol Des* **16**: 825–830.

Sprague PW (1995). Automated chemical hypothesis generation and database searching with Catalyst. *Perspect Drug Disc Des* **3**: 1–20.

Sprague PW, Hoffman R (1997). CATALYST pharmacophore models and their utility as queries for searching 3D databases. In: van de Waterbeemd H, Testa B, Folkers G (eds). *Computer-Assisted Lead Finding and Optimization*. Verlag Helvetica Chimica Acta: Basel, pp 225–240.

Stahura FL, Bajorath J (2004). Virtual screening methods that complement HTS. *Comb Chem High Throughput Screen* **7**: 259–269.

Steindl TM, Schuster D, Laggner C, Langer T (2006). Parallel screening: a novel concept in pharmacophore modeling and virtual screening. *J Chem Inf Model* **46**: 2146–2157.

Stiefl N, Baumann K (2003). Mapping property distributions of molecular surfaces: algorithm and evaluation of a novel 3D quantitative structure–activity relationship technique. *J Med Chem* **46**: 1390–1407.

Stoner CL, Gifford E, Stankovic C, Lepsy CS, Brodfuehrer J, Prasad JV *et al.* (2004). Implementation of an ADME enabling selection and visualization tool for drug discovery. *J Pharm Sci* **93**: 1131–1141.

Sung M-H, Simon R (2004). *In silico* simulation of inhibitor drug effects on nuclear factor-kB pathway dynamics. *Mol Pharmacol* **66**: 70–75.

Swaan PW, Ekins S (2005). Reengineering the pharmaceutical industry by crash-testing molecules. *Drug Disc Today* **10**: 1191–1200.

Testa B, Balmat AL, Long A, Judson PN (2005). Predicting drug metabolism – an evaluation of the expert system METEOR. *Chem Biodiver* **2**: 872–885.

Testa B, Krämer SD (2006). The biochemistry of drug metabolism – an introduction. Part 1: principles and overview. *Chem Biodivers* **3**: 1053–1101.

Todeschini R, Consonni V (2000). *Handbook of Molecular Descriptors*. Wiley-VCH: Weinheim.

Toledo-Sherman LM, Chen D (2002). High-throughput virtual screening for drug discovery in parallel. *Curr Opin Drug Discov Dev* **5**: 414–421.

Verma RP, Hansch C (2005). An approach toward the problem of outliers in QSAR. *Bioorg Med Chem* **13**: 4597–4621.

Vigers GP, Rizzi JP (2004). Multiple active site corrections for docking and virtual screening. *J Med Chem* **47**: 80–89.

Walters WP, Goldman BB (2005). Feature selection in quantitative structure-activity relationships. *Curr Opin Drug Discov Dev* **8**: 329–333.

Wang S, Zaharevitz DW, Sharma R, Marquez VE, Lewin NE, Du L *et al.* (1994). The discovery of novel, structurally diverse protein kinase C agonists through computer 3D-database pharmacophore search. Molecular modeling studies. *J Med Chem* **37**: 4479–4489.

Wermuth CG, Ganellin CRLP, Mitscher LA (1998). Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1997). *Annu Rep Med Chem* **33**: 385–395.

Whittaker PA (2003). What is the relevance of bioinformatics to pharmacology? *Trends Pharmacol Sci* **24**: 434–439.

Wilkens SJ, Janes J, Su AI (2005). HierS: hierarchical scaffold clustering using topological chemical graphs. *J Med Chem* **48**: 3182–3193.

Willett P (2003). Similarity-based approaches to virtual screening. *Biochem Soc Trans* **31**: 603–606.

Willett P, Barnard JM, Downs GM (1998). Chemical similarity searching. *J Chem Inf Comp Sci* **38**: 983–996.

Wolber G, Langer T (2005). LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J Chem Inf Model* **45**: 160–169.

Xue L, Bajorath J (2000). Molecular descriptors in chemoinformatics, computational combinatorial chemistry, and virtual screening. *Comb Chem High Throughput Screen* **3**: 363–372.

Yamashita F, Itoh T, Hara H, Hashida M (2006). Visualization of large-scale aqueous solubility data using a novel hierarchical data visualization technique. *J Chem Inf Model* **46**: 1054–1059.

Yan SF, King FJ, He Y, Caldwell JS, Zhou Y (2006). Learning from the data: mining of large high-throughput screening databases. *J Chem Inf Model* **46**: 2381–2395.

Zhang Q, Muegge I (2006). Scaffold hopping through virtual screening using 2D and 3D similarity descriptors: ranking, voting, and consensus scoring. *J Med Chem* **49**: 1536–1548.