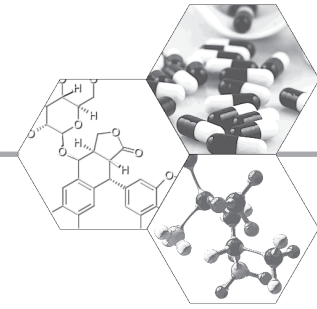

Computational tools for polypharmacology and repurposing

Most drugs act on a multitude of targets rather than on one single target. Polypharmacology, an upcoming branch of pharmaceutical science, deals with the recognition of these off-target activities of small chemical compounds. Due to the high amount of data to be processed, application of computational methods is indispensable in this area. This review summarizes the most important *in silico* approaches for polypharmacology. The described methods comprise network pharmacology, machine learning techniques and chemogenomic approaches. The use of these methods for drug repurposing as a branch of drug discovery and development is discussed. Furthermore, a broad range of prospective applications is summarized to give the reader an overview of possibilities and limitations of the described techniques.

For many decades the 'one drug–one target–one disease' approach was a predominant principle in drug discovery and development. Numerous clinical studies on approved drugs demonstrate that this assumption is a highly restricted view on the action of drugs. The opposite is true: the most active pharmaceutical ingredients interact with several targets in pharmacologically relevant concentrations [1]. This fact is not necessarily malicious but opens numerous chances for pharmaceutical research. On the one hand, the efficacy of a drug might be due to its multitarget profile, whilst on the other hand such dirty drugs offer opportunities for repurposing for new indications. **Polypharmacology** is the upcoming branch of pharmaceutical science dealing with these phenomena. The objective of polypharmacology is the recognition of off-target activities of a small chemical compound. There is a high interference between polypharmacology and chemogenomics. Chemogenomics describes the relationship between targets relating structure and activity of their ligands [2]. The information about ligands of one target and its distance in biological space to other targets can be used to support the evaluation of new molecules for one or even more novel targets. The knowledge gained from both approaches can be used in the early-development stages to filter out compounds and minimize the risk of failure due to severe side effects. Applied to known drugs, polypharmacological methods can lead to repurposing of a compound to a different indication [3]. Marketed drugs or development candidates that have failed in clinical trials, due to lack of efficacy, exhibiting a good safety profile and pharmacokinetic properties are ideal candidates for **drug repurposing**. A classic example for repositioning during development is Sildenafil, which has been in clinical trials for cardiovascular indication and was approved for treatment of erectile dysfunction, a side effect turned into main indication. As stated the repurposing approach can be also applied to already marketed drugs as demonstrated by the example of Memantine [4], which had been marketed previously and was further developed for treatment of moderate to severe Alzheimer's disease.

An experimental technique to detect polypharmacology is the so-called target-fishing approach. Here, a ligand is immobilized on a solid phase bead, such as agarose, and gets exposed to the protein mixture of interest. Targets that bind to the ligand will be later identified using MS or immunoblotting. The main drawback of this method is the unspecific binding of proteins to the immobilized chemical compound. Furthermore, the different levels of proteins in a cell prevent the straightforward application of this method. Therefore, the application of *in silico* methods for this challenge is needed.

From the point of view of a computational chemist the recognition of polypharmacology, as well as repurposing, can be regarded as the matching of a chemical compound with a multitude of targets. This stands in contrast to virtual screening, where a multitude of chemical compounds is matched with one target. Although applications of established virtual-screening techniques could be successfully adopted for polypharmacological purposes, the differences in the general task justify the independent

Janosch Achenbach[1], Pekka Tiikkainen[2], Lutz Franke[2] & Ewgenij Proschak[†1]

[1]Institute of Pharmaceutical Chemistry, Goethe-Universität Frankfurt a. M., ZAFES/LIFF/OSF, Max-von-Laue Str. 9, 60348 Frankfurt, Germany
[2]Chemical R&D, Merz Pharmaceuticals GmbH, Eckenheimer Landstrasse 100, 60318 Frankfurt, Germany
[†]Author for correspondence:
E-mail: proschak@pharmchem.uni-frankfurt.de

FUTURE SCIENCE part of fsg

Key Terms

Polypharmacology: Enumeration of macromolecule targets of a small molecule.

Drug repurposing: Discovery of new indications for already marketed drugs.

development of *in silico* techniques. This article gives an overview of established approaches and successful applications of *in silico* techniques for repurposing and polypharmacology.

Methods

Highly diverse methods have been employed to perform the mapping of chemical compounds to a whole range of targets or indications. In general, we should differentiate between ligand-based and target-based approaches. The essential assumption postulated by Johnson *et al.*, which is also relevant for virtual screening, is that similar compounds exhibit similar biological activity [5]. For detection of polypharmacology, it has to be extended, such as similar targets can interact with similar ligands. Comparable to virtual screening, the definition of similarity is highly dependent on the system. In many virtual-screening campaigns a combination of a specific descriptor and similarity measurement give the best results for a specific target. For detection of polypharmacology, more general models are required, due to the fact that interaction with a great variety of different systems is required.

For the detection of polypharmacology we have a great amount of data at our disposal. This data includes basic biochemical measurements, cellular assay results and *in vivo* activities. Several databases provide an excellent fundament for the application of *in silico* approaches. We do not intend to give a complete overview of different information resources for polypharmacology in this publication, however two public resources are worth mentioning. DrugBank summarizes the data on nearly 4600 approved drugs and experimental compounds in clinical studies [6]. ChemblDB is a database with more than 600,000 distinct compounds with almost 3,000,000 measured biological activities [101]. A successful *in silico* tool for detection of polypharmacology should be able to deal with such amounts of data. High computational speed and an intuitive representation of the results are, hence, crucial criteria for the development of new methods. Therefore, it is not surprising that data-mining and machine-learning techniques widely applied in computer science are convenient for this task.

■ Self-organizing maps

Self-organizing maps (SOMs) belong to the category of unsupervised machine learning techniques and can be used in various application areas of medicinal chemistry, such as design of screening library, scaffold-hopping and drug repurposing [7]. SOMs are based on the principle that a set of connected neuron vectors are used to approximate the distribution of the input space by assigning similar input data points to the same and most similar neuron (weight vector). Therefore, SOMs allow projecting high-dimensional input data on a lower dimension (often 2D) given by an intrinsic dimensionality of the neuronal network **(Figure 1)**.

The training consists of two major steps that are iterated in each training epoch. In the competitive step a random input vector will be chosen and presented to the whole neuron network to determine the so-called 'winner neuron', the weight vector with the highest similarity to the input vector. Afterwards, the adaptive step updates all weight vectors resulting in a movement of all neurons toward the winner neuron. The movement range depends on the distance between each neuron and the winner neuron defined by a neighborhood function and the network topology. Due to the adaptive learning rate the network converges.

Self-organizing maps can be used as a ligand-based approach to find new off-targets for a given set of compounds. Therefore, a SOM will be trained on a set of molecules with known biological activity and than a set of chemical compounds of interest will be projected on this map. Afterwards, all annotated compounds used for training, which were assigned to the same neurons as the projected molecules, are analyzed relating their targets.

■ Bayesian classifiers

Naive Bayesian classifier is a popular supervised data-mining technique often employed in computer science. It is based on the assumption that every feature (in the case of small molecules, a pharmacophore or a substructure) depends only on the class it belongs to (bioactivity on a certain target) and not on other features. Although this is not true in the most cases, naive Bayesian classifiers are successful in practice. Nidhi *et al.* used chemical descriptors [8], more precisely extended-connectivity fingerprints with multiple-category naive Bayesian classifier trained on compounds of 964 targets of the World of Molecular Bioactivity (WOMBAT) chemogenetic database [9]. The extended-connectivity fingerprints are circular substructural fingerprints based on the Morgan Algorithm, which features represent particular structures with restricted and definite attachment points [10].

The authors trained a model on 85% of the WOMBAT compounds to predict the targets of the remaining 15% for model evaluation. The model guessed the correct target for 92% and the correct target family for 95% of the compounds among the top three predictions.

Afterwards the MDL Drug Data Report (MDDR) database, (licensed by Molecular Design, Ltd, San Leandro, CA, USA [102]) was exposed to the trained model and showed for ten different MDDR activity classes an average correct target prediction in the top three estimates for 77% of the compounds and for 79% of the compounds the correct target family. Furthermore, they were able to relate WOMBAT compounds with only generic activity annotations, such as 'kinase inhibitor' or 'antineoplastic inhibitor', to individual MDDR targets with described therapeutic activity.

The target-based model was extended to consider, in addition, protein domains. This approach attempts to bypass the problem that target-based models can only predict targets, for which annotated ligands are contained in the training sets. In order to provide a generalization they expanded their model by assuming that similar ligands not only bind to the same target but also to receptors having similar amino acid sequences or protein folds.

■ Similarity ensemble approach algorithm

In contrast to model-based approaches, the similarity ensemble approach (SEA) algorithm offers a model-free alternative [11]. The method has been used to identify pharmacological links between molecular targets based on set similarities of their respective ligands. Since only active molecules are being compared, no inactive decoy molecules need to be picked as has to be done with model-based tools. In addition, there is no learning step required before the prospective prediction.

The approach goes beyond simple mean pair-wise similarity calculations (i.e., reporting the average similarity of all ligand pairs of two sets) by adding a statistical element to the calculations. The sum of similarities (raw score) greater than a predetermined threshold between molecules in the two sets is calculated. The raw score alone has a bias for product of sizes of the two sets and, therefore, is corrected by the mean and standard deviation of raw scores expected had the two sets consisted of randomly picked molecules. Finally, the algorithm produces an expectation value for the pair **(FIGURE 2)**.

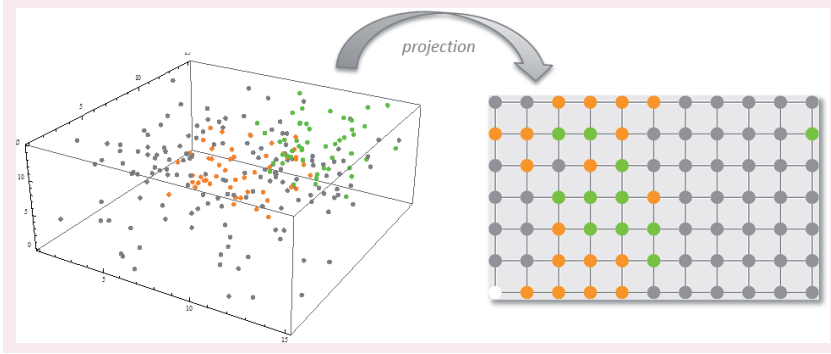

Figure 1. Projection of 3D data on a 2D self-organizing map.

Keiser *et al.* applied the methodology to compare ligand sets from a subset of the MDDR database [11]. Their data set was comprised of 65,241 ligands split into 246 sets, each assigned to a distinct molecular target. On average each target was linked to 5.8 other targets with an E-value $\leq 10^{-10}$ and with only 18 targets that did not link with any other target. Out of the 246 targets, 193 could be linked to a protein sequence, which were aligned and scored pairwise. Comparison of the biological and chemical similarities of the targets yielded some surprising results. For example, serotonin receptors were not linked to opioid receptors on the chemical level despite their high sequence similarity. The target classes are very similar outside their ligand-binding regions but lack sufficient active site similarity to bind the same ligands. However, the authors do not discuss the option of restricting sequence alignment to active sites only. Perhaps this would have led to results more consistent with the ligand set similarities.

The SEA algorithm has been successfully applied to drug repurposing. These results are detailed in the Applications section of this article.

■ Network pharmacology

In recent years, the advantages of network and systems biology have become more important in drug discovery, especially in drug repurposing and polypharmacology. Based on the theory that complex diseases arise from a change of the equilibrium of a system, cell or protein complex (the healthy state) we need drugs that shift the equilibrium in order to minimize this change [12]. To develop drugs that meet those standards it is essential to understand the underlying system, the interactions between different parts, such as proteins, metabolites, RNA and other cellular entities, and the effect of perturbing several parts of that system by drugs.

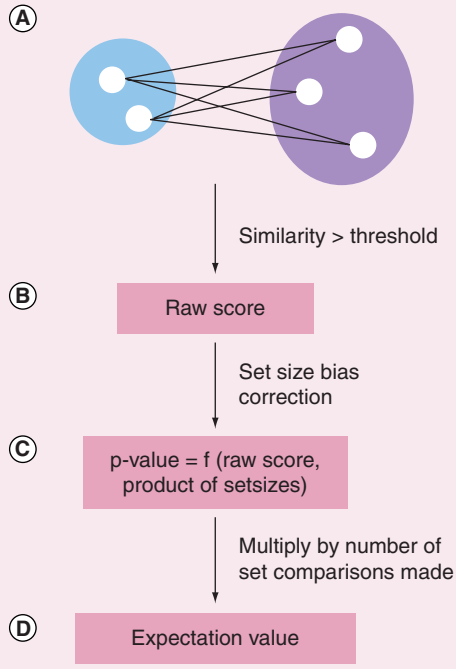

Figure 2. Schematic overview of the similarity ensemble approach algorithm. **(A)** All pair-wise similarities are calculated for ligands in the two sets being compared and **(B)** those exceeding a predefined threshold are summed to come up with the raw score for the set pair. **(C)** The raw score is normalized into a p-value by comparing the raw score to values expected if the ligand sets had been randomly picked (random background function). **(D)** Finally the p-value is multiplied by the total number of set comparisons made.

A so-called drug–target network represents this system by relating information of drugs, targets and their interactions achieved by genomics, metabolomics, transcriptomics and proteomics **(FIGURE 3)**. These networks can in addition be partitioned into drug–drug and target–target networks. The former network type describes the relations between drugs sharing a significant number of targets, while target–target networks relate targets, which are targeted by the same set of drugs. In addition, information gained by structure–activity data for compounds that bind multiple targets can be used for a better understanding of polypharmacology [2]. Network pharmacology can be further used to design drugs that are improved in clinical efficacy based on a better understanding of side effects and toxicology. Furthermore, the understanding of complex diseases on a molecular level linked to the physiological context of the related targets leads to disease-associated networks. These networks offer new therapeutic views and, in addition, recommendations for new indications of existing drugs [13]. Consequently, the next step is the active design of multi-target drugs that affect several parts of a complex biological system to cure complex diseases [14].

Excessive knowledge on complex pathways allows the simulation of metabolic and signaling pathways, or even whole cellular environments.

The computational techniques developed in this area are applicable for detection of potential combination of targets for efficient therapy but also for simulation of polypharmacological agents. Dynamic simulations of a metabolic network can be performed using a series of ordinary differential equations. Each metabolic reaction is represented by differential Michaelis–Menten equation. The equation parameters are often obtained from Kyoto Encyclopedia of Genes and Genomes [15]. Solver routines are used to fit the ordinary differential equation system to experimental. The obtained model can be used to observe the influence of multi-target inhibitors [16] or drug combinations [17].

■ Pharmacophores

Many ligand-based approaches rely on descriptors that compare molecules on a very concrete level of the atom connectivity, which limits the structural diversity of the results. Pharmacophores have been used for a long time to overcome this caveat [18]. Interactions between the ligand and the target are represented on an abstract level as pharmacophore features without an explicit requirement for a hit molecule to have a certain atom connectivity. This property makes pharmacophores directly applicable for unearthing ligand–target links not obvious from atom connectivity alone. Pharmacophores can be generated by overlaying a set of ligands and identifying shared pharmacophore features. If a co-crystallized protein model exists, the key interactions between the ligand and the protein can be directly defined to derive a structure-based pharmacophore.

■ Binding site similarity

Milletti and Vulpetti introduced a method to predict polypharmacology by pocket similarity [19]. This approach is based on two steps, including the calculation of pocket shape similarity and a subsequent alignment for refinement. They use the method to set up so-called inhibition maps describing the compound promiscuity to several targets.

In the first step, they take n points of the pocket contour and describe the shape context of each point by counting the occurrences of other points within 13 spheres of increasing radii, r. These points represent atoms typed by their physical chemical properties.

N fingerprints define the whole pocket, each for every typed atom, describing the coarse distribution of the pocket shape in its local context. Consequently the pocket similarity is calculated by matching point pairs with same 'shape context' in two distinct pockets.

For the final screening procedure the authors define the query pocket 3 Å around the bound ligand. Then similarity scores for each combination of atoms of the query pocket and the pockets in the database are calculated and stored in a matrix, K.

These scores are used to find the best atom equivalences of both pockets, which can be formally seen as a linear assignment problem [20]. It can be solved by calculating permutations over the matrix K to find a minimum score S, which is the sum of all atom pair scores.

The obtained list of atom equivalences is employed to align the pockets using a modified Procrustes algorithm [21]. Then the previous Score, S, is refined with regard to this alignment by multiplying with the superimposed atoms root mean-square deviation, normalized by the query pocket size and scaled from 0 to 100 (optimal match).

Applications

The SEA algorithm has been applied in prospective drug repurposing and previously unknown drug–target associations were found. In the first paper where the algorithm was introduced, methadone was predicted and experimentally confirmed to antagonize M3 muscarinic receptor [11]. In addition, over 12,000 compounds in PubChem were screened and the drugs emetine and loperamide were experimentally validated to antagonize adrenergic α2 and neurokinin NK2 receptors, respectively.

In another paper by the Shoichet group, SEA found more previously unknown associations involving 15 drugs and 23 targets [22]. Most of the predictions involved known aminergic G-protein coupled receptor (GPCR) binders linked to other targets of the same protein family. More interestingly, there were also four drugs crossing target class boundaries. For example, the HIV-1 reverse transcriptase inhibitor Rescriptor was found to bind the histamine H_4 receptor (GPCR).

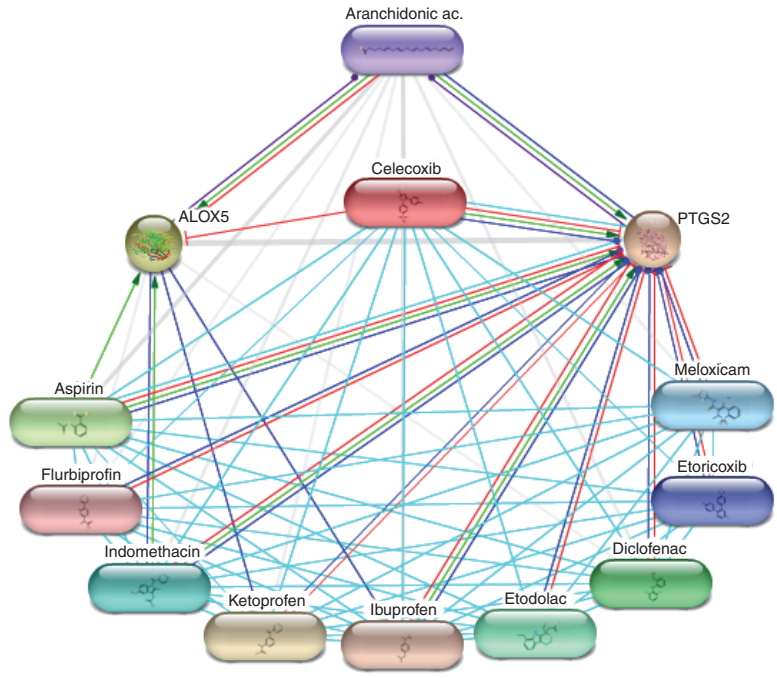

Figure 3. Drug–target network created with Stitch 2 [35], describing the relationships between 5-lipoxigenase, cyclooxygenase-2, their common substrate arachidonic acid and a set of inhibitors that binds to one of both targets as well as both targets such as celecoxib. Modes of action are shown in different colors.

Markt *et al.* built a set of 48 peroxisome-proliferator activated receptor (PPAR) pharmacophore models from protein crystal structures co-crystallized with various ligands [23]. After the automatic extraction, the models were further manually refined to fit the training set of known ligands. Out of the 48 models, the three best performing plus one ligand-based model were chosen for the final step. Here, a set of 321 known PPAR activators and 36 inactive molecules were screened against a panel of pharmacophore models representing 181 protein targets. The known actives were found to rank a PPAR model in the top ten 81% of the time whereas 60% of the inactives did so. The relatively small difference between actives and inactives plus a small number of the latter would warrant more studies to assess the applicability of the models in prospective screens. In addition, the authors did not show how many of the test set actives 'finding' a PPAR model were actually structural analogs of the ligand used to build the model and, therefore, would have been found with more simple 2D methods as well.

A publicly available web service called PharmMapper is available online, which allows the user to screen through more than 7000 pharmacophore models [24], built using the same

algorithm as used by Markt *et al.* In the accompanying paper, only a very limited evaluation of the tool using tamoxifen is presented making it impossible to rate the usefulness of the tool at the moment this article was written.

Schneider *et al.* used self-organizing maps in a polypharmacology context by projecting aspirin on a SOM trained on approved drugs and leads [25]. Self-organizing maps were used to predict a new target for aspirin, tyrosine phosphatase 1B. Noeske *et al.* also used self-organizing maps to find new targets for known antagonists of metabotrobic glutamate receptor subtypes 1 and 5 [26]. Their SOM predicted other GPCRs, such as dopamine D_2 and D_3, muscarinic acetylcholine receptors and histamine H_1 as possible targets, which were experimentally verified with IC_{50} values between 5 and 100 µM.

Milletti and Vulpetti used the binding site similarity approach to predict inhibition maps for 17 kinase inhibitors in regard to their polypharmacology profile and were able to predict targets exhibiting K_d < 10 µM even with low sequence identity between the targets [19]. An advanced application demonstrated the usefulness of this approach for all targets with structural information stored in the Protein Data Bank for a set of ten known biologically active compounds and showed a variety of protein families, which could be associated with the active compounds by binding site similarity analysis. Both proteins with similar and different functions were retrieved. One of the used queries were levonorgestrel bound to the progesterone receptor, which was related with the androgen receptor, the mineralocorticoid receptor, glucocorticoid receptor and estrogen receptor, which are all known targets for Levonogestrel. In the case of sitaglipin as a query, a selective dipeptidyl peptidase (DPP)4 inhibitor, the approach found seprase as an a potential related target, which is beside of DPP8 and DPP9 a known target of many other DPP4 inhibitors.

Gregori-Puigjané and Mestres applied successfully Drug–Target Networks on data obtained from an *in silico* target profiling of 767 drugs on 684 targets [27]. The network showed that some aminergic GPCRs were highly connected to NMDA, sigma and opioid receptors, despite any phylogenetic relationships, which was confirmed by annotated examples from literature.

The work of Scheiber *et al.* describes an analysis of adverse drug reactions (ADRs) linked to chemical features that are correlated to particular effects [28]. They used a data set of marketed drugs and their annotated ADRs to build a map of ADRs relationships. Therefore, they extracted 4210 different ADR terms from the Medical Dictionary of Regulatory Activities and their associated molecules. The molecules were described by extended-connectivity fingerprint descriptors and further used for the training of multiple-category Laplacian-modified naive Bayesian classification models for each ADR term. Then each model is paired and the similarity is defined by the Pearson correlation between their normalized feature probabilities. Pairs with an correlation of $\rho > 0.8$ are retained as significant and were used to build an ADR interaction network with Cytoscape [29]. They were able to correlate different ADRs of drugs by mapping common chemical features proved by examples from literature.

Bisson *et al.* demonstrated an *in silico* drug repurposing approach for the identification of novel nonsteroidal androgen receptor antagonists [30]. The authors used glucocorticoid receptor and estrogen receptor crystal structures in an antagonist-bound conformation as a template to establish preliminary models of the androgen receptor ligand-binding domain stabilized by a binding antagonist. These were further iteratively optimized by docking with nuclear receptor ligands and a Monte Carlo energy minimization procedure and yielded two structural variant models. These models were used in a docking study of marketed oral drugs and led to 11 compounds tested in a competitive binding assay. Three of them containing a phenothiazine scaffold showed low-micromolar K_i values in competition to [3H] mibolerone binding. The drugs are used in the treatment of schizophrenia and inhibit the dopamine D2 receptor and the $5HT_2$ family of the serotonin receptors. Known endocrine side effects of these drugs include loss of sexual desire and impotence, which could be explained by the low adrenergic receptor antagonistic activity. The phenothiazine system, which is topologically similar to the rigid steroid scaffold of the natural ligands, was used for a substructure search and a set of ten phenothiazine derivatives were screened for their antiandrogen activity, where the most potent inhibitor an IC_{50} potency of 25 µM.

Discussion

The current increase in drug repurposing activities in most pharmaceutical companies and specialized service providers is a direct response to the current productivity crisis with the need to reduce the attrition rates in drug development. As, especially, large pharmaceutical companies have rich resources of failed drug candidates, specialized

divisions have been founded and cooperation agreements signed. Part of the effort also led to an increase in development and application of *in silico* methods in this domain.

The presented studies demonstrate the applicability of *in silico* methods for polypharmacology analysis and drug repurposing. Although already successfully applied, there is a lot of space for improvement and further development in the coming years.

The methods used in the past have mainly relied on 2D representation of small molecules due to computational limitations. First, 3D methods have been already described but further development will offer possibilities for unearthing target–target associations not possible when limiting oneself to the 2D world. Which, taken together with recent advances in computational throughput of 3D tools (e.g., developments in graphics processing unit computing [31]), give promise that these methods can be applied on the same scale as 2D tools today.

Although the focus of the current article has been on chemoinformatic tools, we should note that small molecules can also be compared computationally based on the phenotypes they elicit. These include indication specific *in vivo* models, drug side effects [32] and gene-expression profiles [33]. Given the different approach these tools offer, we consider these as complementary and possibly even synergetic with the methods described in this article – a possibility that warrants research in the future.

Owing to the fact that the efficacy of many approved drugs is based on the activity on more than one target, conscious and rational design of polypharmacologically active compounds should bring a change in drug discovery as a logical consequence. The demand for this paradigm change has been stated previously [14,34] and recent advances in *in silico* polypharmacology could be the stimulating factor for its realization.

Future perspective

The idea of drug repurposing has been followed since the first drugs were developed. But only in recent years there have been intensified efforts by pharmaceutical companies, specialized service providers and academia to systematically identify novel indications for drugs already known or discontinued in clinical development.

In silico methodologies have advanced as a valued technique in early drug discovery and as more and more target structures, structure bioactivity data and, therefore, optimized chemoinformatic tools come available they are likely to expand there impact within that process. As drugs with a specific polypharmacologic profile will permit a superior treatment of certain diseases, one of the computational key tasks lying ahead will be the implementation and development of methodologies able to identify relevant compounds.

In regard to this, and as it makes so much economic sense to perceive the opportunity of shortened drug-development time, risk and costs, *in silico* predictions and drug repositioning will stay a growing area of research within the years to come.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties. No writing assistance was utilized in the production of this manuscript.

Executive summary

- Promiscuity of drugs lead to the development of polypharmacology, a novel branch of pharmaceutical science.
- Off-target effects of small molecules can be used for repurposing of a chemical entity to a novel indication. Computational methods provide valuable information in this area.
- A large variety of algorithms have been developed for recognition and prediction of polypharmacological behavior of small molecules. The *in silico* techniques summarized in this review comprise network pharmacology, machine learning techniques and chemogenomic approaches as main categories:
  - Network pharmacology relates various pharmacological data (binding affinities, *in vitro* data, *in vivo* studies and side effects) to pharmacological compounds and allows the exploration of potential drug–target interactions;
  - Machine learning is suitable to recognize similarity patterns and is used for prediction of off-target activities of small chemical compounds;
  - Chemogenomic approaches are used to study the interactions of whole compound classes with families of related targets.
- Several applications of the mentioned approaches are described in the second section of the review indicating the benefits of computational approaches in polypharmacological studies.

Bibliography

1 Sams-Dodd F. Target-based drug discovery: is something wrong? *Drug Discov. Today*. 10(2), 139–147 (2005).

2 Paolini GV, Shapland RHB, van Hoorn WP, Mason JS, Hopkins AL. Global mapping of pharmacological space. *Nature Biotechnol*. 24(7), 805–815 (2006).

3 Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. *Nature Rev. Drug Discov*. 3(8), 673–683 (2004).

4 Reisberg B, Doody R, Stöffler A, Schmitt F, Ferris S, Möbius HJ. Memantine in moderate-to-severe Alzheimer's disease. *N. Engl. J. Med.* 348(14), S1333–S1341 (2003).

5 Johnson M, Lajiness M, Maggiora G. Molecular similarity: a basis for designing drug screening programs. *Prog. Clin. Biol. Res.* 291, 167–171 (1989).

6 Wishart DS, Knox C, Guo AC *et al.* DrugBank: a comprehensive resource for *in silico* drug discovery and exploration. *Nucleic Acids Res.* 34(Database issue), D668–D672 (2006).

7 Kohonen T. Self-organized formation of topologically correct feature maps. *Biol. Cybern*. 43, 59–69 (1982).

8 Nidhi, Glick M, Davies JW, Jenkins JL. Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. *J. Chem. Inf. Model*. 46(3), 1124–1133 (2006).

9 Olah M, Mracec M, Ostopovici L *et al.* WOMBAT: world of molecular bioactivity. In: *Chemoinformatics in Drug Discovery*. Oprea TI (Ed.). Wiley-VCH, Weinheim, Germany 221–239 (2004).

10 Morgan HL. The generation of a unique machine description for chemical structures – a technique developed at chemical abstracts service. *J. Chem. Doc.* 5(2), 107–113 (1965).

11 Keiser MJ, Roth BL, Armbruster BN *et al.* Relating protein pharmacology by ligand chemistry. *Nature Biotechnol*. 25(2), 197–206 (2007).

12 Chandra Janga S, Tzakos A. Structure and organization of drug–target networks: insights from genomic approaches for drug discovery. *Mol. Biosyst*. 5, 1536–1548 (2009).

13 Pujol A, Mosca R, Farrés J, Aloy P. Unveiling the role of network and systems biology in drug discovery. *Trends Pharmacol. Sci*. 31(3), 115–123 (2010).

14 Hopkins AL. Network pharmacology: the next paradigm in drug discovery. *Nature Chem. Biol.* 4(11), 682–690 (2008).

15 Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res*. 38(Database issue), D355–D360 (2010).

16 Yang K, Ma W, Liang H *et al.* Dynamic simulations on the arachidonic acid metabolic network. *PLoS Comput. Biol*. 3(3), 8 (2007).

17 Iadevaia S, Lu Y, Morales FC, Mills GB, Ram PT. Identification of optimal drug combinations targeting cellular networks: integrating phospho-proteomics and computational network analysis. *Cancer Res*. 70(17), 6704–6714 (2010).

18 Langer T. Pharmacophores in drug research. *Mol. Inf.* 29(6–7), 470–475 (2010).

19 Milletti F, Vulpetti A. Predicting polypharmacology by binding site similarity: from kinases to the protein universe. *J. Chem. Inf. Model.* 50(8), 1418–1431 (2010).

20 Jonker R, Volgenant A. A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Computing* 38(4), 325–340 (1987).

21 Rohlf FJ, Slice DE. Extensions of the procrustes method for the optimal superimposition of landmarks. *Syst. Zool.* 39(1), 40–59 (1990).

22 Keiser MJ, Setola V, Irwin JJ *et al.* Predicting new molecular targets for known drugs. *Nature* 462(7270), 175–181 (2009).

23 Markt P, Schuster D, Kirchmair J, Laggner C, Langer T. Pharmacophore modeling and parallel screening for PPAR ligands. *J. Comp. Aided Mol. Des.* 21(10–11), 575–590 (2007).

24 Liu X, Ouyang S, Yu B *et al.* PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Res*. 38(Web server issue), W609–W614 (2010).

25 Schneider P, Tanrikulu Y, Schneider G. Self-organizing maps in drug discovery: compound library design, scaffold-hopping, repurposing. *Curr. Med. Chem*. 16(3), 258–266 (2009).

26 Noeske T, Jirgensons A, Starchenkovs I *et al.* Virtual screening for selective allosteric mGluR1 antagonists and structure–activity relationship investigations for coumarine derivatives. *ChemMedChem* 2(12), 1763–1773 (2007).

27 Gregori-Puigjané E, Mestres J. A ligand-based approach to mining the chemogenomic space of drugs. *Comb. Chem. High Throughput Screen.* 11(8), 669–676 (2008).

28 Scheiber J, Jenkins JL, Sukuru SCK *et al.* Mapping adverse drug reactions in chemical space. *J. Med. Chem.* 52(9), 3103–3107 (2009).

29 Shannon P, Markiel A, Ozier O *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 13(11), 2498–2504 (2003).

30 Bisson WH, Cheltsov AV, Bruey-Sedano N *et al.* Discovery of antiandrogen activity of nonsteroidal scaffolds of marketed drugs. *Proc. Natl Acad. Sci. USA* 104(29), 11927–11932 (2007).

31 Giupponi G, Harvey MJ, de Fabritiis G. The impact of accelerator processors for high-throughput molecular modeling and simulation. *Drug Discov. Today*. 13(23–24), 1052–1058 (2008).

32 Campillos M, Kuhn M, Gavin A-C, Jensen LJ, Bork P. Drug target identification using side-effect similarity. *Science* 321(5886), 263–266 (2008).

33 Iorio F, Bosotti R, Scacheri E *et al.* Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc. Natl Acad. Sci. USA* 107(33) 14621–14626 (2010).

34 Morphy R, Rankovic Z. Designing multiple ligands – medicinal chemistry strategies and challenges. *Curr. Pharm. Des.* 15, 587–600 (2009).

35 Kuhn M, Szklarczyk D, Franceschini A *et al.* STITCH 2: an interaction network database for small molecules and proteins. *Nucleic Acids Res.* 38(Database issue), D552–D556 (2010).

■ Websites

101 European Bioinformatics Institute. ChemblDB www.ebi.ac.uk/chembl

102 Molecular Design, Ltd www.mdli.com