Research Article   **Open Access**

Computational Strategies for Drug Reprofiling

Serban San-Marina[1,2]*, Rajeesh Gupta[2] and Ionel Iosif[1]

[1]*University of Toronto, BiostatistiX, Canada*
[2]*University of Toronto, Canada*
[3]*BiostatistiX, Canada*

Abstract

This article introduces some of the recent developments in drug re-profiling with emphasis on how computational chemistry and biology approaches together with access to public data bases can help generate new leads from existing drugs. It discusses the drawbacks of high-throughput (HTS) genomics and how the concepts of target poly-pharmacology can help speed up delivery of the next generation of drugs. Some of the successful strategies for drug re-profiling are presented and computational tools are discussed.

The problem with HTS-Genomics

Completion of the Human Genome Project promised to dramatically speed up the discovery of new medicines. Knowledge of the complete sequence of all protein coding genes meant that molecules could now be designed to target specific amino acid sequences. At about the same time, advances in combinatorial chemistry made it possible to generate chemical libraries of increasing complexity containing small, drug-like molecule. Combined with the pioneering successes in molecular biology over the previous two decades, it is now possible to synthesize any protein and pan an entire combinatorial library against it in the hopes of finding specific, high affinity leads for the next generation of drugs. When genomics and robotics joined forces genomics high-throughput screening (genomics-HTS) was born and quickly became the main thrust for lead development. A variety of drugs were produced using genomics-HTS including imatinib mesylate developed for the treatment of chronic myelocytic leukemia by Brian Drucker at the University of Oregon. Gleevec dramatically changed the course of a disease with grim prognosis for thousands of people world-wide. However, as the number of New Drug Application (NDA) submissions to the US Food and Drug Administration has remained flat since the introduction of genomics-HTS relative to the period prior to it these early successes did not translate into increased productivity. The problem is not with the generation of sufficient well-targeted leads, but with the high attrition rates of these leads, particularly in the clinic. Specifically, these targeted molecules repeatedly violate Lipinsky's rule of five requirements, for example by coming in at higher than optimal molecular sizes [1]. The failure of HTS-genomics to significantly increase productivity is only adding to the well-documented drug pipeline crunch. The quadrupling of R&D expenditures over the last 25 years and the heavy reliance on revenues generated by a handful of 'block-buster' drugs with only a few years left in their patents threaten to dry out pharma pipelines in the near future. With insufficient funds coming in will pharma afford the hefty R&D bill for new medicines? Leading researchers suggest that a new generation of selectively promiscuous drugs could solve the pipeline crisis [2]. Because of shorter pipeline transit times and lower development costs re-profiling promiscuous old drugs for new uses is an attractive alternative to the traditional new chemical entity development.

The re-profiling paradigm

Roll back the clock by 20 years and the idea of drug re-profiling was every bit of a pharmacological heresy. Back then the distinction was clear between the desirable or therapeutic effects of drugs and their undesirable side effects which were thought to be 'non-specific'. Therapeutic effects result when drug molecules interfere with the activity of specific protein targets, often through binding with good affinity to amino acid sequences delineating three-dimensional cavities or pockets in enzymes. In the absence of a good pharmacological explanation, pharmacokinetics was blamed for the fact that not all patients with the same condition benefited from the same drug. The status of the intended targets, i.e. wild-type or mutated was not considered. Of substantially more concern was the fact that a percentage of all patients receiving the same medication experienced adverse side effects. Unfortunately, side effects were usually coaxed in such broad terms that in general, they did not suggest the involvement of separate mechanisms and specific targets. It is difficult to pinpoint with certainty when the realization hit that some of the side effects can be as 'specific' as the primary therapeutic effects and can create opportunities for re-profiling. The re-profiling of Viagra is often quoted as an example of how a `bust` was turned into a `block-buster` but one of the earliest examples of successful re-profiling was the re-branding of the anti-hypertensive drug Minoxidil for the treatment of male-pattern baldness. From retinoic acid being successfully re-profiled for the treatment pro-myelocytic leukemia, to the use of thalidomide in the treatment of non-Hodgkin's lymphoma, the rush to re-profile is on.

Re-profiling: Plan A

More recently, formalised approaches to drug re-profiling have been initiated using computational tools developed to assess drug-drug and target-target likeness. By analysing existing targets and ligands it was thought that formal principles can be deduced to help predict the nature and number of all potential targets and thus determine the size of the druggable genome. Prior to sequencing the human genome up to 10,000 drug targets were predicted but a post-sequencing 2002 study set the number closer to ~3,000 [3]. By contrast, a recent study puts the number of confirmed drug targets at only 218 [4]. These predictions

were based on the amino acid sequences of the three most highly represented targets used in drug development, namely G protein-coupled receptors, kinases and nuclear receptors and on the similarities between the common drug matter that binds to these proteins. In a different approach taken by Han and colleagues a support vector machines algorithm was used to map the physico-chemical features of druggable targets, rather than their amino acid sequences and to derive a list of predicted targets that are compliant with these features [5]. This study sets the size of the druggable genome at 3379. What makes these studies possible is the observation that similar proteins bind closely related ligands and similar ligands bind to a constant set of targets. Ligand similarity is routinely determined by the Tanimoto coefficient which is the ratio of the common and the distinct features of any two ligands and takes values between 0 (no similarity) and 1 (identity). The SuperDrug and SuperLigand databases also allow 3D superimpositions of ligands to determine which of the conserved side groups are implicated in interactions with a given target. By using protein-protein and ligand-ligand similarity searches, the existing drug-target (DT) pharmacological space has been mapped. It displays every interaction between all known drugs and targets. One remarkable feature of the DT landscape is the extent of drug promiscuity. For example, the aminergic GPCR family of D(2) dopamine receptors bind over 8,000 active compounds, SRC kinases almost 1,800 and Protein Kinase C delta type almost 200. Paolini and colleagues provide a simplified visual reference guide to the DT landscape with additional information available by request from these authors [6] while Yamanishi and colleagues used the KEGG database to create their own map of the DT landscape [7]. Spreadsheets linking drugs to multiple targets and targets to multiple drugs are available from these authors by means of a limited license. The STITCH database is also freely available to the public without registration and it provides a graphical interface that links multiple drugs and targets. Another useful resource is the ID Map, a freely downloadable Java application where MDDR and ASINEX libraries data on ~600,000 compounds have been linked with assay bioactivity data. This tool offers access to more than just FDA approved drugs but does not link compounds to individual targets.

Network poly-pharmacology

Although these resources link drugs to multiple targets, it can be a daunting task to identify diseases in which the activity of the targets is up-regulated. With the possible exception of some of the more important kinases, activity data is generally not available for proteins across the wide spectrum of human diseases. Fortunately, extensive experience with gene expression profile analysis over the last decade shows that in general, gene expression data can be used as a surrogate for protein activity. Nevertheless, targeting just one gene product may not bring about the expected therapeutic benefit given that diseases are rarely caused by a single aberrantly expressed protein and more often involve a network of interacting proteins. Thus, what is further required is some knowledge of network pharmacology. Cellular networks of interacting proteins consist of hubs, which are hotspots receiving multiple inputs, nodes which are individual non-hub proteins and vertices that link hubs and nodes. Each hub is characterized by its degree which refers to the total number of vertices that connect to it. The so-called bottleneck hubs funnel the flow of network information through a single connection to the hub of an adjacent network, while non-bottleneck hubs have multiple connections to neighbouring networks. Disrupting bottleneck hubs is the best strategy for disrupting

a network. Having decided how to select which type of networks to target for disruption how can one find disease networks to use for drug re-profiling? To construct a network one would need to download experimental microarray data. The NIH's Gene Expression Omnibus (GEO) and the EMBL-EBI's Array Express contain the majority of all publically available microarray data. Free gene expression analysis software can be obtained from The Institute for Genomic Research or similar organizations. The goal of the gene expression analysis is to identify the set of up-regulated targets which will contain the "druggable' genes. Once these genes are identified, programs such as HiMap or STITCH can be accessed free of charge to build the network of interacting genes.

Re-profiling: Plan B

Another way of using gene expression data for drug re-profiling is to convert the information contained in microarray experiments into a set of "standards" that will allow navigation of the genomic landscape for the drug to be re-profiled by testing it against other existing profiles. Profiles available in public databases include FDA-approved drugs, other investigational compounds, knock-down siRNA experiments as well as chronic diseases including most of the common malignancies. Tools such as EXALT allow to test a query signature against all signatures deposited in the NIH GEO public data base. Another excellent tool is Cmap2 at the Broad Institute which compares an uploaded profile to over 7,000 other profiles generated using FDA-approved as well as some other chemicals commonly used in cellular experimentation. The value of these tools consists in the fact that the drug to be re-profiled is evaluated in terms of how similar it is to other drugs, whose mechanisms of action are well understood and for which the targets are well defined. In some cases however, the price paid for being able to query this massive volume of microarray data is that some of the matches may not be particularly informative. For example, it is difficult to interpret the significance of a match between the drug of interest and a subset of samples exhibiting a clinical condition or an experimental alteration, such as hypoxia. Therefore, whenever possible it is a good idea to compile a proprietary library of expression profiles against which to pan the expression data for the re-profiled drug, provided of course that a suitable algorithm to enable the comparison is first selected. In short, similar gene expression profiles suggest similar mechanisms of action and similar targets. Likewise, if the gene expression data for the re-profiled drug matches the signature of a knock-down experiment, the knocked down protein may constitute a valid target for the drug. Using an algorithm based on correlation analysis has the added advantage that positive as well as negative correlations can be established. For example, if the re-profiled drug is tested against a panel of human diseases, a high negative correlation value (closer to -1) would suggest that the drug might be able to reverse some of the symptoms associated with the disease.

Conclusion

Drugs helped us discover how the living body works and in turn this knowledge has made the next generation of drugs more specific and less dangerous. Each cycle of re-invention also absorbed the leading scientific ideas of the time. Bioinformatics, computational chemistry, network poly-pharmacology and drug promiscuity concepts could transform drug research to the point that twenty years from now a scientist looking back at drug development might just wonder how "one gene, one drug, one disease" has dominated so much of twentieth century pharmacology.

References

1. Mencher SK, Wang LG (2005) Promiscuous drugs compared to selective drugs (promiscuity can be a virtue). BMC Clin Pharmacol 5:3.

2. Hopkins AL (2008) Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol 4:682-690.

3. Hopkins AL, Groom CR (2002) The druggable genome. Nat Rev Drug Discov 1:727-730.

4. Imming P, Sinning C, Meyer A (2006) Drugs, their targets and the nature and number of drug targets. Nat Rev Drug Discov 5:821-834.

5. Han LY, Zheng CJ, Xie B, Jia J, Ma XH, et al. (2007) Support vector machines approach for predicting druggable proteins: recent progress in its exploration and investigation of its usefulness. Drug Discov Today 12:304-313.

6. Paolini GV, Shapland RH, van Hoorn WP, Mason JS, Hopkins AL (2006) Global mapping of pharmacological space. Nat Biotechnol 24 :805-815.

7. Yamanishi Y, Araki M, Gutteridge A, Honda W, Kanehisa M (2008) Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. Bioinformatics 24:i232-i240.