Bioorganic & Medicinal Chemistry 20 (2012) 5461–5463

Contents lists available at SciVerse ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

Backtranslating clinical knowledge for use in cheminformatics—What is the potential?

Josef Scheiber *

BioVariance, Garmischer Str. 4/V, 80339 Munich, Germany

ARTICLE INFO

Article history:
Available online 4 May 2012

Keywords:
Cheminformatics
Translational research
Personalized medicine
Clinical data

ABSTRACT

'From bench to bedside' is seeing a very strong focus in current Drug Discovery. However, often overlooked are the advantages that turn out if data is used 'from bedside to bench', the fact one can also make beneficial use of clinical information in early Drug Discovery. By leveraging the wealth of clinical data carried by each marketed drug, down to the level of a single person, one can gain a deep insight that can be leveraged in conjunction with chemical structure information and therefore within all kinds of cheminformatics analyses. This supports the design of drugs that better fit the requirements of a well-defined subpopulation.

Within this contribution I am going to focus on the realm of cheminformatics applications and how this data can thereby used to better impact the decisions of medicinal chemists.

Recent Drug Discovery is seeing a very strong focus on the 'from bench to bedside' approach, leveraging research information to optimize clinical treatment decisions. Themed as translational research in drug discovery and development, it typically refers to the translation of non-human research findings, from the laboratory and from animal studies, into therapies for patients.

However, often overlooked are the advantages that turn out if data are used 'from bedside to bench,' and the fact that one can also make beneficial use of clinical information in early Drug Discovery. Each marketed drug carries a wealth of clinical data that can be leveraged in conjunction with chemical structure information, and therefore within all kinds of cheminformatics analyses. For example, structure information can be used to link fragments of newly synthesized compounds with knowledge about detrimental as well as beneficial effects from marketed drugs to then design drugs that better fit the requirements of a certain subpopulation. Detrimental here is defined as knowledge that supports the understanding of unwanted effects from marketed drugs, for example, what off-targets are hit because of a certain substructure whereas beneficial alludes to knowledge about wanted activity, that is, for example reducing the activity or abundance of a target linked to a symptom of the treated disease.

These data can then be used in all stages of early Drug Discovery (i.e., target finding, target validation), but in this contribution, I am going to focus on the realm of cheminformatics applications and how this data can influence analyses consecutively used to impact the decisions of medicinal chemists.

This requires a significant change in the present situations. What are the points that need to be considered if one wants to fully leverage the potential of using patient data as input for analyses like Lead Finding or Lead Optimization? Everyone is familiar with filtering out undesirable known-chemical-toxic fragments and substructures. However, using late stage clinical and post marketing data to enhance cheminformatics results is a quite novel idea, bringing for example, filters based on clinical outcomes into place that will avoid a certain unwanted side effect or off-target interaction later on. One can think this through a few steps further. Following news of drug approvals, the current trend towards personalized medicine comes to mind. Figure 1 illustrates the idea in an overview. Designing tailored drugs specifically for a set of patients going further than a classical repositioning into a new indication requires adoption of certain strategic measures to understand the biology of a human subpopulation in more detail. This is possible and is not that tough if there is effective use of the already available vast amount of knowledge from earlier projects aiming towards a similar effect or knowledge that can be quickly generated at constantly falling prices. For example, genomic knowledge within electronic health records providing phenotypic information are a hot trend at the moment. The myriads of open source knowledge bases both in the biology and chemical area provide many more good examples.

Though clinical data usage is gaining more acceptances there are several hurdles in this path. Various challenges involve cultural, technological, standardization, protocol driven science, work flow usability and validation for studies. The process is resource intensive as well and includes various stages like collecting of data, checking and analyzing as well as a final presentation.[1] Moreover,

* Tel.: +49 89 1896582 80; fax: +49 89 1896582 99.
E-mail addresses: mail@josef-scheiber.de, josef.scheiber@biovariance.com

0968-0896/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.bmc.2012.04.054

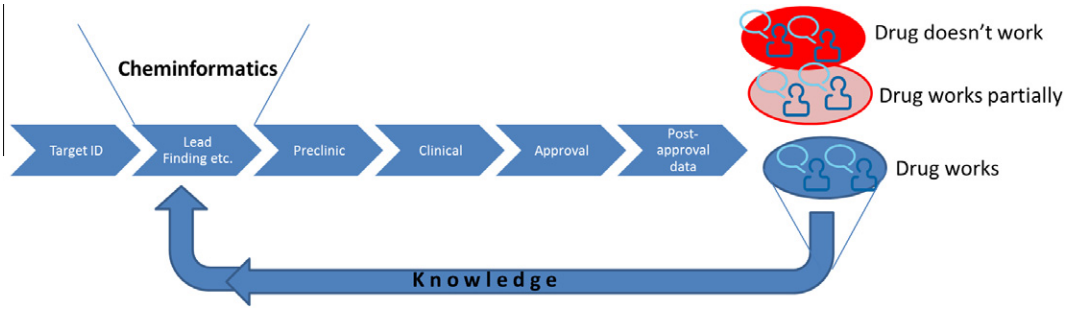

Figure 1. Here the position of cheminformatics within the Drug Discovery Pipeline is shown. Whereas the usual focus is on Lead Finding driven by data from earlier stages in the future one will also be able to use clinical information to focus the analysis on more relevant facts where the developed drugs work as intended. The pipeline will be closer to a knowledge circle.

the whole process beginning from defining of data that is to be collected to presenting it necessitates utilization of sophisticated technology by professionals who are highly skilled to then really enable an analysis in chemical context. However, for the sake of this manuscript I consider these solvable challenges as a non-issue for the future and will therefore not dive into detail.

Though there are prerequisites,[2] the hypothesis is supported by two recent examples: The first is Plexxikon's Vemurafenib.[3] It interrupts the step from B-Raf to MEK on the pathway of B-Raf/MEK/ERK. Vemurafenib only works if B-Raf has the common V600E mutation (glutamic acid replaces normal valine and position number of amino acid is 600 on B-Raf). Melanoma patients having this mutation are the only ones affected by vemurafenib. Vemurafenib will not inhibit melanoma cells that are not having this particular mutation.[4]

Another, even more impressive example is Vertex' Ivacaftor.[5] It works for a particular subset of cystic fibrosis patients by targeting the G551D (Aspartic acid replaces the normal Glycine at position 551) mutation, found in only 4% of patients in the CFTR gene.[6] Chloride ions cross the cell membranes through the channel created by CFTR protein. This is a vital stage in the production of sweat, digestive enzymes and mucus. When a patient is having the G551D mutation the CFTR protein fails to open this channel effectively and ions fails to pass through properly. The latter problem is remedied by Ivacaftor. In both cases, there is a clear link between the patient's genetic make-ups to the mode of action of a drug.

These are very specific examples and the results cannot be immediately translated into the development of other drugs. However, the point that I want to make is, that there will be many more equally interesting and detailed datasets available with such a deep understanding of patient genetics and even proteomics in the near future.[7] Inevitably supported by the current data tsunami from next generation sequencing this will lead to the identification of many more similar examples, which will have a clear impact on cheminformatics: The knowledge that is used as input for analysis will be vastly different from what we are used to have thus far. The clinical data will be less noisy than nowadays due to a better patient stratification, it will be a lot easier to understand patient-specific differences and therefore the targets and off-targets themselves. By focusing and enhancing the information a better starting point is generated for cheminformatics analyses: They can and will be run with narrower, better-defined datasets, enabling results that point much more clearly to possible changes on a molecule that is supposed to drive the desired effect on the patient's biology. When trying to find a lead for a well-defined patient subset with a defined mutation or epigenetic modification one has much more detailed information available about the target under scrutiny. The recommendations communicated to medicinal chemists will be much more to the point and much better supported by actual knowledge. This will give a much better credibility to the analysis outcome and therefore reduce the black-box-feeling non-experts nowadays often have when confronted with results. Another point is that one can leverage information into me-too drugs, which are very similar to an existing drug but will be having some minor differences, that actually act a lot better than the initially approved version. Actually, more cheminformatics expertise will be required as analyses for the same target in different versions in different subpopulations have to be run.

The scale at which sequence data, and therefore valuable knowledge about targets is being generated, is growing with an impressive and constantly increasing pace. As a consequence, the current hype around next generation sequencing technologies will heavily drive clinical readouts to a very detailed level, and a lot of this data will also become available for researchers in an anonymized version. Along the same lines, phenotypical knowledge about the patient is being collected: How does the drug work for the patient? What problems are there? What dose is needed? What symptoms are there? Which other drugs are taken? Most of these data points will in the future be collected in a structured manner within Electronic Health records. This means that one can combine this valuable data with already available chemical databases and run large scale analyses on top to elucidate how certain chemical fragments can be linked with all the other descriptions to then steer compound optimizations.

Recent years have seen several publications in which approaches that aim to use clinical information in earlier stages to lead discovery projects have been presented. Potential of clinical information in early Drug Discovery has been found stated in several publications recently. Particularly, information about clinical side effects has been investigated intensively in order to better understand the underlying chemistry and biology. Personalized medicine, pharmacogenomics and next generation sequencing are interrelated. This is because a truly personalized medication is possible only if the patients genome is sequenced to then understand the pharmacogenomics that deals with genetical influence of the patient's metabolism in response to drugs.[8–15] Also, Chemogenomics knowledge has successfully been used to find additional indications for drugs;[16,17] however, it might also be beneficial to initially restrict newly found indications for a drug based on patient data to increase the rate at which a drug will actually work for the patients that receive it, and therefore gain approval for a limited patient subset. One can capitalize on all these approaches by bringing clinical and genetic knowledge into play to then steer medicinal chemistry developments.

The clinical data can in addition to knowledge about the target itself also be linked to other available facts. For example, this also will lead to a better understanding of Pharmacogenomics as com-

pound-target interactions can be linked to specifically modified targets which will help to better understand drug metabolism related issues on patient level. Both the toxicity properties and off target interactions can then be investigated by pulling additional pieces of information together with Chemogenomics databases like ChEMBL by using well-defined identifiers that make sure facts about the same entity are taken from each database. Thereby a link from the biological activity against a certain target to a side effect can be established, enabling medicinal chemists to work on that specific property. The combination of cheminformatics analysis results with supporting knowledge can then be presented to a medicinal chemist in a very concise manner because the supporting data points can in many cases be linked to established drugs that are understood very well and therefore deliver very welcome input. Consequently, putting the data tsunami from genetics together with knowledge-based approaches that are already available or being continuously developed in Cheminformatics will have a significant impact on the results of the analyses. Such results can be much more exact, due to the better and more available genetical sequencing data for input, which beneficially focus possible outcomes of analyses.

In summary, extending the horizon of cheminformatics by incorporating other information types will inevitably create a very big playground for smart technologies and also expand the current use cases. An exciting use case is for example, to envisage approaches where chemical similarity can be used to identify possible biomarkers and thereby drive the development of companion diagnostic tests. This can be accomplished by linking changes in gene expression profiles to certain chemical substructures. If the same changes occur in molecules with similar structure one will already have a very strong hint, driven by cheminformatics.

Ultimately, approaches linking data from early Drug Discovery projects and Clinical use to put them into use in early Drug Discovery will turn out to be extremely beneficial in the upcoming era of Personalized Healthcare. This will then lead to the fact that Cheminformaticians will be seen as a much more serious partner by medicinal chemists, as they will be able to support their own results with much better and broader supporting knowledge. There is bright future ahead for Cheminformatics if one opens up the mind towards additional data types.

Acknowledgment

I am very grateful for the very interesting and challenging input provided by four expert referees that helped to significantly improve an earlier version of this manuscript.

References and notes

1. Lu, Z.; Su, J. *J. Clin. Trials* **2010**, *1*, 1–2.
2. Scheiber, J. *Expert Opin. Drug Discov.* **2011**, *3*, 1–6.
3. Bollag, G.; Hirth, P.; Tsai, J., et al *Nature* **2010**, *467*, 596–599.
4. Sala, E.; Mologni, L.; Truffa, S., et al *Mol. Cancer Res.* **2008**, *6*, 751–759.
5. Ledford, H. *Nature* **2012**, *482*, 145.
6. Ramsey, B. W.; Davies, J.; McElvaney, N. G., et al *N. Eng. J. Med.* **2011**, *365*, 1663–1672.
7. Chen, R.; Mias, G. I.; Li-Pook-Than, J., et al *Cell* **2012**, *148*, 1293.
8. Bender, A.; Scheiber, J.; Glick, M., et al *ChemMedChem* **2007**, *2*, 861.
9. Scheiber, J.; Chen, B.; Milik, M., et al *J. Chem. Inf. Model.* **2009**, *49*, 308.
10. (a) Scheiber, J.; Jenkins, J. L.; Sukuru, S. C. K., et al *J. Med. Chem.* **2009**, *52*, 3103; (b) Yang, L.; Agarwal, P. *PLoS One* **2011**, *6*, e28025. Epub 2011 Dec 21.
11. Luo, H.; Chen, J.; Shi, L., et al *Nucleic Acids Res.* **2011**, *39*, W492.
12. Yang, L.; Wang, K.; Chen, J., et al *PLoS Comput. Biol.* **2011**, *7*, e1002016. Epub 2011 Mar 31.
13. Ekins, S.; Nikolsky, Y.; Bugrim, A.; Kirillov, E. *Methods Mol. Biol.* **2007**, *356*, 319. Review.
14. Ekins, S.; Andreyev, S.; Ryabov, A., et al *Drug Metab. Dispos.* **2006**, *34*, 495. Epub 2005 Dec 28.
15. Ekins, S.; Nikolsky, Y.; Nikolskaya, T. *Trends Pharmacol. Sci.* **2005**, *26*, 202. Review.
16. DeGraw, A. J.; Keiser, M. J.; Ochocki, J. D., et al *J. Med. Chem.* **2010**, *53*, 2464. Epub 2009 Nov 1.
17. Keiser, M. J.; Setola, V.; Irwin, J. J., et al *Nature* **2009**, *462*, 175. Epub 2009 Nov 1.