Development and application of computational methods for the identification and optimization of bioactive compounds

Johannes Kirchmair
Center for Bioinformatics, University of Hamburg, Bundesstrasse 43, 20146 Hamburg, Germany
kirchmair@zbh.uni-hamburg.de

In September 2014 we started building up a new research lab for applied cheminformatics and molecular design at the Center for Bioinformatics of the University of Hamburg. Our major research interests are the development and application of computational methods for the identification and optimization of bioactive compounds, in particular for drug metabolism prediction, target prediction, virtual screening, conformer ensemble generation, and natural product research. Here we provide a brief summary of our recent activities in two of these research areas.

1 Development of computational methods for drug metabolism prediction

Metabolism of small organic molecules can yield metabolites with substantially different physicochemical and biological properties [Kirchmair 2013a]. Consequently, understanding biotransformation is of immediate relevance to the safety and efficacy of drugs, cosmetics, nutritional supplements and agrochemicals. Today a plethora of experimental methods are available which allow the generation of a fairly complete picture of the metabolic fate of small molecules but they remain expensive and time-consuming. Driven by these factors as well as ethical issues related to the use of animal models, computational methods for drug metabolism prediction have become an active field of research [Olsen 2015, Peach 2012, Kirchmair 2015, Kirchmair 2014, Kirchmair 2012].

The primary bottleneck for computational tools is the scarcity of high-quality data on drug metabolism. The largest database on metabolic reactions, Metabolite (BIOVIA, San Diego, CA), contains about 100k substrate-metabolite pairs organised in about 14k metabolic pathways. It covers about 90% of all approved drugs but only a small fraction of drug-like molecules, natural products and human endogenous metabolites.

We have developed data mining methods for the automated analysis of metabolic reactions and pathways such as the ones stored in Metabolite. In a first study we used these techniques to analyse how and to which extent the metabolic system changes the physicochemical properties of small organic molecules (including drugs, endogenous metabolites, and molecules related to traditional Chinese medicine) [Kirchmair 2013a]. For example, we could show that drug metabolism produces metabolites with a calculated logP that is on average one log unit lower than that of the parent compound. Interestingly, this shift toward more hydrophilic molecules is much less pronounced for endogenous metabolites such as nutrients and micronutrients, which are retained in the body. Such methods allowed us to identify specific metabolic reactions and enzymes, which, against the global trend, result in more lipophilic metabolites. This knowledge e.g. can be applied to the design of skin care products with a prolonged retention time at the target tissue.

In a second study we developed a random forest-based predictor of sites of metabolism (regioselectivity): FAME (FAst MEtabolizer) [Kirchmair 2013b]. The sites of metabolism of about 20k molecules of Metabolite were automatically annotated using the MetaPrint2D software framework [Adams 2010]. Seven 2D chemical descriptors encoding the element, hybridisation state, electronic properties and steric accessibility were calculated for all atoms of all molecules. A collection of random forest models was then trained on subsets of this data. Individual models were computed for human, rat, dog and mammalian metabolism. Reaction classification was used to derive dedicated models for phase 1 and phase 2 metabolism. FAME correctly identifies at least one known site of metabolism among the top-

1, top-2, and top-3 highest-ranked atom positions in up to 71%, 81%, and 87% of all cases tested, respectively. These success rates are comparable to or better than other models focused on specific enzyme families (such as cytochrome P450s; CYPs), yet FAME covers a very broad chemical space (drugs, endogenous metabolites and natural products) and a fairly comprehensive set of reactions and enzymes relevant to xenobiotic metabolism. In a complementary approach we used three probabilistic machine learning methods, Parzen-Rosenblatt Window (PRW), Naive Bayesian (NB) and RASCAL (Random Attribute Subsampling Classification ALgorithm) for the generation of highly accurate models for site of metabolism prediction [Tyzack 2014]. The classifiers were implemented in CUDA/C++ for GPU acceleration and obtained top-2 success rates of about 80-90% for CYPs 3A4, 2D6 and 2D9.

A plateauing in prediction accuracy of methods for site of metabolism prediction is observed, and the primary reason for this appears to be the limitations of current datasets with respect to coverage of the chemical space, diversity, completeness, correct assignment of sites of metabolism and stereochemical information. Here, substantial efforts in data collection and curation are to be made. Future research directions will include the implementation of more advanced descriptors for a more accurate representation of the chemical reactivity and steric accessibility of atoms.

2 Computer-guided identification and optimization of bioactive compounds

Currently our research group is actively pursuing a dozen national and international research collaborations with experimentalists on the identification and optimization of bioactive compounds. Much of our recent drug discovery projects have been focussed on viral [Richter 2015, Grienke 2011, Kirchmair 2011, Grienke 2010, von Grafenstein 2015] and bacterial [von Grafenstein 2015, Walther 2015] neuraminidases, which are scientifically highly interesting to address using computational techniques because of their pronounced conformational flexibility and specific structural properties. For example, influenza virus neuraminidase is only active when in a quaternary assembly, but neuraminidases of other biological systems, such as bacteria, are active as monomers. Using molecular dynamics simulation techniques we derived a hypothesis of the structural basis of this assembly dependency [von Grafenstein 2015]. Understanding this specific requirement of influenza neuraminidases is of immediate relevance to the structure-based design of new inhibitors, which so far has often relied on structures derived from simulations of the monomer of the viral enzyme. As we know now, simulation of this system is most likely insufficient because of the significant impact of the assembly state on the conformation of the active site.

We successfully applied a shape-based screening method to identify a variety of plant constituents from *Glycyrrhiza glabra* [Grienke 2011] and others [Kirchmair 2011] as inhibitors of influenza neuraminidase. A shape-based screening method also allowed us to identify benzylhydantoin and related compounds as *in vivo* highly effective chemical chaperons of phenylalanine hydroxylase [Santos-Sierra 2012]. These compounds can be used to treat phenylketonuria, an inherited deficiency caused by protein misfolding. Current treatment options for this disease are very limited, costly and often not effective.

Recently we identified inhibitors of the interaction of protein kinase C (PKC) epsilon and RACK2 [Rechfeld 2015] using a pharmacophore model. PKCepsilon has been related to neoplastic transformation, cardiac hypertrophy, nociceptor function and others. The model was derived from the structure of the C2 domain of PKCepsilon and used to screen a commercial library of 330k molecules for potential disruptors. Nineteen compounds were purchased and tested in *in vitro* assays. One of the tested compounds (based on a thienoquinoline scaffold) showed moderate activity as a disruptor of this protein-protein interaction, and the best out of 19 analogues tested in a follow-up study, N-(3-acetylphenyl)-9-amino-2,3-dihydro-1,4-dioxino[2,3-g]thieno[2,3-b]quinoline-8-carboxamide, had an IC_{50} of 5.9 micromolar (which is comparable to that of a dodecapeptide fragment of RACK2 binding to this protein-protein interface).

References

[Adams 2010] Adams, S. E. Molecular similarity and xenobiotic metabolism. Ph.D. Thesis. University of Cambridge, UK. 2010.

[Grienke 2011] Grienke, U.; Braun, H.; Seidel, N.; Kirchmair, J.; Richter, M.; Krumbholz, A.; von Grafenstein, S.; Liedl, K. R.; Schmidtke, M.; Rollinger, J. M. Computer-guided approach to access the anti-influenza activity of licorice constituents. *Journal of Natural Products*, 2014, 77, 563-570.

[Grienke 2010] Grienke, U.; Schmidtke, M.; Kirchmair, J.; Pfarr, K.; Wutzler, P.; Drrwald, R.; Wolber, G.; Liedl, K. R.; Stuppner, H.; Rollinger, J. M. Antiviral potential and molecular insight into neuraminidase inhibiting diarylheptanoids from *Alpinia katsumadai*. *Journal of Medicinal Chemistry*, 2010, 53, 778-786.

[Kirchmair 2015] Kirchmair, J.; Gller, A. H.; Lang, D.; Kunze, J.; Testa, B.; Wilson, I. D.; Glen, R. C.; Schneider, G. Predicting drug metabolism: Experiment and/or computation? *Nature Reviews Drug Discovery*, 2015, 14, 387-404.

[Kirchmair 2014] Kirchmair, J. (ed.) Drug Metabolism Prediction Wiley: Weinheim, 2014. ISBN 978-3-527-33566-4

[Kirchmair 2013a] Kirchmair, J.; Howlett, A.; Peironcely, J.; Murrell, D. S.; Williamson, M. J.; Adams, S. E.; Hankemeier, T.; van Buren, L.; Duchateau, G.; Klaffke, W.; Glen, R. C. How do metabolites differ from their parent molecules and how are they excreted? *Journal of Chemical Information and Modeling*, 2013, 53, 354-367.

[Kirchmair 2013b] Kirchmair, J.; Williamson, M. J.; Afzal, A. M.; Tyzack, J. D.; Choy, A. P. K.; Howlett, A.; Rydberg, P.; Glen, R. C. FAst MEtabolizer (FAME): A rapid and accurate predictor of sites of metabolism in multiple species by endogenous enzymes. *Journal of Chemical Information and Modeling*, 2013, 53, 2896-2907.

[Kirchmair 2012] Kirchmair, J.; Williamson, M. J.; Tyzack, J. D.; Tan, L.; Bond, P. J.; Bender, A.; Glen, R. C. Computational prediction of metabolism: Sites, products, SAR, P450 enzyme dynamics and mechanisms. *Journal of Chemical Information and Modeling*, 2012, 52, 617-648.

[Kirchmair 2011] Kirchmair, J.; Rollinger, J. M.; Liedl, K. R.; Seidel, N.; Krumbholz, A.; Schmidtke, M. Novel neuraminidase inhibitors: Identification, biological evaluation and investigations of the binding mode. *Future Medicinal Chemistry*, 2011, 3, 437-450.

[Olsen 2015] Olsen, L.; Oostenbrink, C.; Jrgensen, F. S. Prediction of cytochrome P450 mediated metabolism. *Advanced Drug Delivery Reviews*, 2015, doi: 10.1016/j.addr.2015.04.020.

[Peach 2012] Peach, M. L.; Zakharov, A. V.; Liu, R.; Pugliese, A.; Tawa, G.; Wallqvist, A.; Nicklaus, M.C. Computational tools and resources for metabolism-related property predictions. 1. Overview of publicly available (free and commercial) databases and software. *Future Medicinal Chemistry*, 2012, 4, 1907-1932.

[Rechfeld 2015] Rechfeld, F.; Gruber, P.; Kirchmair, J.; Boehler, M.; Hauser, N.; Hechenberger, G.; Garczarczyk, D.; Lapa, G. B.; Preobrazhenskaya, M. N.; Goekjian, P.; Langer, T.; Hofmann, J. Thienoquinolines as novel disruptors of the PKC/RACK2 protein-protein interaction. *Journal of Medicinal Chemistry*, 2014, 57, 3235-3246.

[Richter 2015] Richter, M.; Schumann, L.; Walther, E.; Hoffmann, A.; Braun, H.; Grienke, U.; Rollinger, J. M.; von Grafenstein, S.; Liedl, K. R.; Kirchmair, J.; Wutzler, P.; Sauerbrei, A.; Schmidtke, M. Complementary assays helping to overcome challenges for identifying neuraminidase inhibitors. *Future Virology*, 2015, 10, 77-88.

[Santos-Sierra 2012] Santos-Sierra, S.; Kirchmair, J.; Perna, A. M.; Rei, D.; Kemter, K.; Rschinger, W.; Glossmann, H.; Gersting, S. W.; Muntau, A. C.; Wolber, G.; Lagler, F. Novel pharmacological chaperones that correct phenylketonuria in mice. *Human Molecular Genetics*, 2012, 21, 1877-1887.

[Tyzack 2014] Tyzack, J. D.; Mussa, H. Y.; Williamson, M. J.; Kirchmair, J.; Glen, R. C. Cytochrome P450 site of metabolism prediction from 2D topological fingerprints using GPU accelerated probabilistic classifiers. *Journal of Cheminformatics*, 2014, 29.

[von Grafenstein 2015] von Grafenstein, S.; Wallnfer, H. G.; Kirchmair, J.; Fuchs, J. E.; Huber, R. G.; Spitzer, G.; Schmidtke, M.; Rollinger, J. M.; Liedl, K. R. Interface dynamics explain assembly dependency of influenza neuraminidase catalytic activity. *Journal of Biomolecular Structure & Dynamics*, 2015, 33, 104-120.

[Walther 2015] Walther, E.; Richter, M.; Xu, Z.; Kramer, C.; von Grafenstein, S.; Kirchmair, J.; Grienke, U.; Rollinger, J. M.; Liedl, K. R.; Slevogt, H.; Sauerbrei, A.; Saluz, H. P.; Pfister, W.; Schmidtke, M. Antipneumococcal activity of neuraminidase inhibiting artocarpin. *International Journal of Medical Microbiology*, 2015, 305, 289-297.