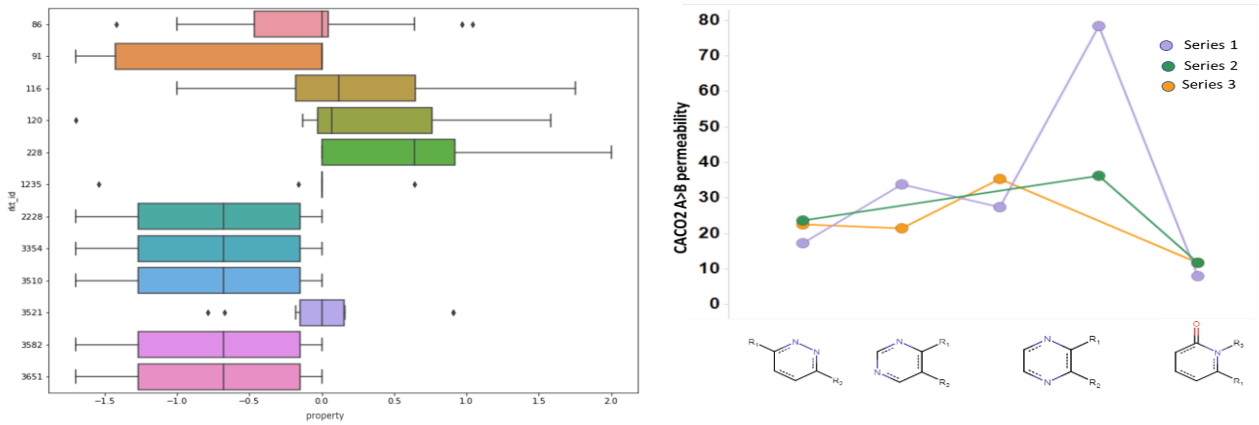
Compound optimization - What we should know about the SAR trends we don't know

Gianna Pohl and Daniel Kuhn

Computational Chemistry & Biology, Merck Healthcare KGaA, Frankfurter Strasse 250, 64293 Darmstadt, Germany



Structure activity relationships (SAR) are important factors in drug development to understand and optimize the physiochemical and biological properties of molecules. An examination of the exchange of individual residual groups can provide information on how these groups influence the properties. Methods to investigate the influence of single residual groups are for instance Matched Molecular Pair (MMP) analysis [1] and Free-Wilson analysis [2]. To perform the necessary calculations of the analyses, there are already freely available programs, such as *mmpdb* [3] or the implementation of the Free-Wilson analysis by Pat Walters [4]. While performing MMP analysis is straightforward, these tools hardly provide any possibilities to visualize the results or to put them into context. This is where our work starts. We created several command line tools to facilitate data preparation, analysis calculations allowing for consideration of different endpoints. *Jupyter notebooks* and *voila* web applications were created to visualize the results in different ways. For example, the distributions of the influence of different transformations can be visualized as a box plot as shown in the figure above left. The results can also be exported to analysis tools such as *Spotfire* or *Jupyter notebooks* to allow analysis by broad scientific audience. These analysis tools allow that activity cliffs can be easily detected, and the influence of certain residual groups can be understood. To analyze trends between related chemical series Matched Molecular Series (MMS) analysis has been implemented. MMP and MMS analysis have been applied to ADME endpoints such as CACO2 permeability and microsomal stability. Analyzing related compound series helps to elucidate SAR trends and provides information how SAR can be transferred across series, e.g. changes in CACO2 permeability are dependent on MMS (see Figure above right). The developed methodology is applied in the optimization of ADME endpoints and examples from drug discovery projects will be shared.

[1] A. G. Dossetter, E. J. Griffen, A. G. Leach, *Drug Discovery Today*, *Matched molecular pair analysis in drug discovery*, **2013**, *18(15-16)*, 724-731
[2] S. M. Free, J. W. Wilson, *Journal of Medicinal Chemistry*, *A Mathematical Contribution to Structure-Activity Studies*, **1964**, *7(4)*, 395-399
[3] A. Dalke, J. Hert, C. Kramer, *Journal of chemical information and modeling*, *mmpdb: An open-source matched molecular pair platform for large multiproperty data sets*, **2018**, *58(5)*, 902-910.
[4] P. Walters, *Free Wilson Analysis*, **2018**, http://practicalcheminformatics.blogspot.com/2018/05/free-wilson-analysis.html