Drug safety assessment through automatic extraction of structure-activity relationships

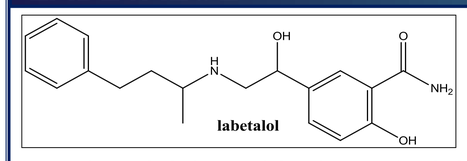
Bernd Wendt, Alexander Steudle

Certara International, Martin-Kollar-Straße 17, 81829 München

INTRODUCTION

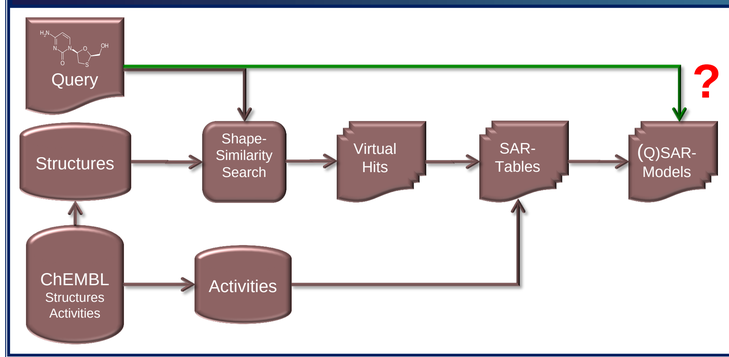
An improved workflow for automatic building of 3D-QSAR models from chemical biology databases such as ChEMBL[1] will be presented. Starting with a chemical structure of choice a 3D-smilarity search identifies neighborhood compounds in chemical biology space. These compounds form the basis for an iterative procedure to produce significant and robust 3D-QSAR models. The resulting models provide indication of potential drug safety threats but also enable a mechanistic understanding of potential toxicity effects by relating the model back to the structure and highlighting those parts of the structure that renders it toxic. Such a drug safety assessment goes beyond similarity and physicochemical property-based computational models and can help the medicinal chemist to make better compounds.

LABETALOL



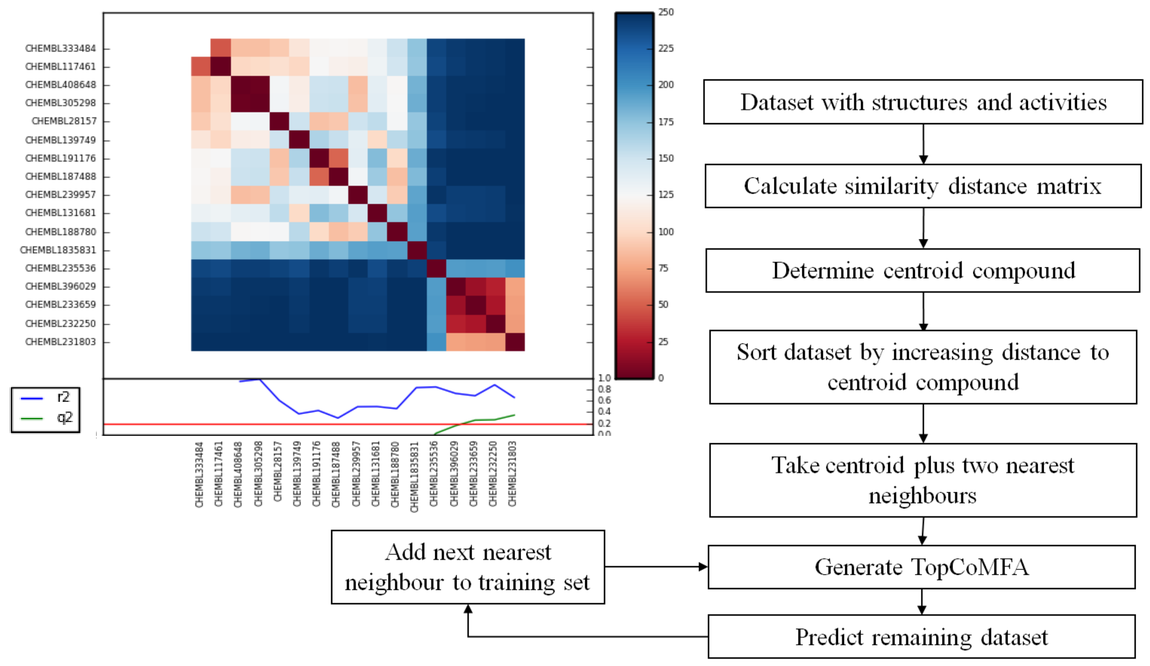
Labetalol is a blocker of alpha- and beta-adrenergic receptors and used as an antihypertensive drug. It can be regarded as a safety promiscuous drug as it acts on many other receptors. Currently it is well known to cause adverse nervous system effects, such as sleep disturbances, a.o.[4]

AUTOMATIC EXTRACTION OF (Q)SARS



SAR tables were constructed for every combination of query structure and target name within the ChEMBL database. This was achieved by mapping the biological activities to the corresponding structures using the database's internal compound identifiers.[3]

QSEA METHOD (Quantitative Series Enrichment Analysis)

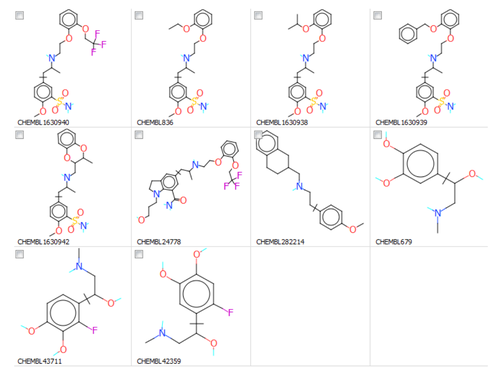


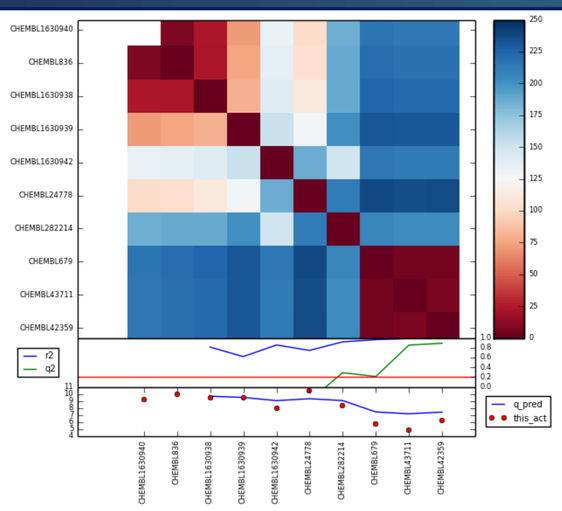
The Topomer CoMFA methodology on which QSEA depends differs from conventional CoMFA only in using topomers as the requisitely aligned 3-D structures. For QSEA, the particular fragment pair then used to represent each molecular structure is automatically chosen as the one providing the lowest mean topomer dissimilarity to the other structures in the SAR table.

QSEA begins with designation of a centroid compound within the data set on the basis of its similarity distance matrix constructed from the topomer descriptor/distance between the compounds. The structure of the centroid compound is then split into all possible two-piece fragmentations using available single bonds. For all resulting two-piece fragmentations, the other structures of the SAR table are checked for their best matching two-piece fragmentation. These sets of corresponding two-piece fragmentations are compared on the basis of their overall mean topomeric distance. The set with the smallest topomeric distance mean is chosen as input for the Topomer CoMFA routine.[2]

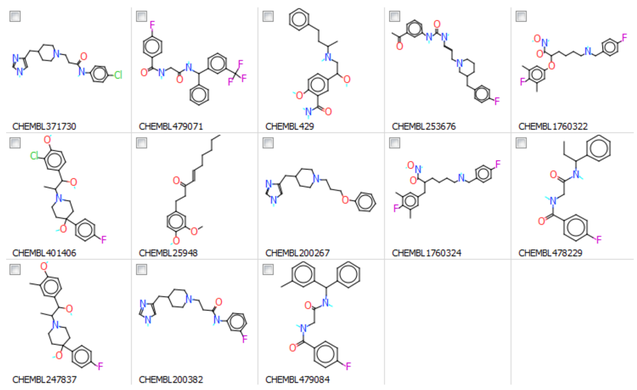
RESULTS

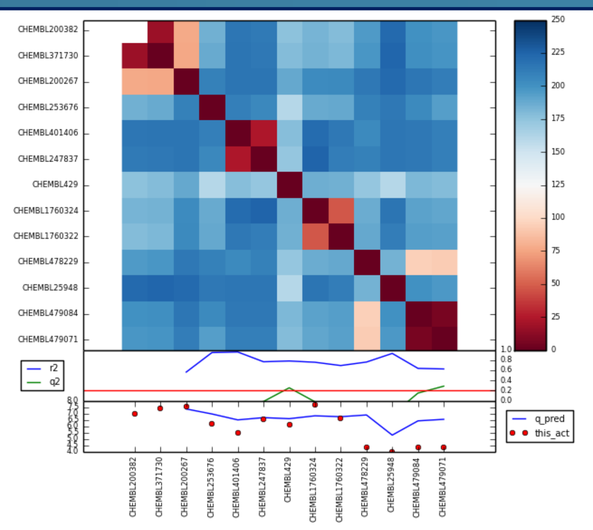
Alpha-1A-Adrenoceptor





CYP450-2D6



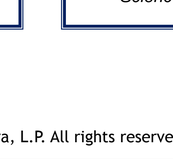


Multi-Target Structure-Activity Relationships:

1. Benzamide fragment

The analysis identified structure-activity relationships for most GPCR-receptors by using ChEMBL data. A selection of SARs including Alpha-1a-Adrenoceptor, Beta-2 Adrenoceptor, Serotonin-1a, and CYP-2D6 are shown.

2. Steric CoMFA contours

Alpha-1a	Beta-2	Serotonin-1a	CYP-2D6
			
3- and 4-position favored	2- and 3-position favored	2- and 3-positions favored	5-position favored
Extended structures in 3-position disfavored	Extended structures in 5-position disfavored	4-position disfavored (in part)	3-position disfavored

3.Electrostatic CoMFA contours

Alpha-1a	Beta-2	Serotonin-1a	CYP-2D6
Heteroatoms favored in 5-position	Hydrocarbons favored in 3- and 4-position	Hydrocarbons favored in 4-position	Hydrocarbons in 3- and 4-position favored
Hydrocarbons favored in 2- and 3-positon	Heteroatoms favored in extended structures of 3- and 4-position	Heteroatoms favored in 2-position	Heteroatoms favored in extended structure in 4-position

Target	Actual	Predicted
Alpha-1a	8.1	7.2
Beta-2	8.6	7.2
Serotonin-1a	6.4	6.1
CYP-2D6	6.1	5.2

- SAR is different for different targets
- Composition of training sets for different SARs is different, or similar
- Composite interpretation of SARs enable design of new leads

REFERENCES

1. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res*. **2001**, *40*, D1100–D1107.
2. Wendt, B.; Cramer RD., Quantitative Series Enrichment Analysis (QSEA), J. Comp. Aid. Mol. Des. 2008, 22, 541-555
3. Wendt, B.; Uhrig, U.; Bös, F. Capturing Structure−Activity Relationships from Chemogenomic Spaces. *J. Chem. Inf. Model.* **2011**, *51*, 843–851.
4. Dahlöf, C.; Dimenäs, E. Side Effects of [beta]- Blocker Treatments as Related to the Central Nervous System. *American Journal of the Medical Sciences,* **1990**, 4, 236-244.