Development of a Hybrid Chemotype Rulebase for *In Silico* Screening of Cholestasis

M. Cronin,[1] C. Yang,[2] S. Belfield,[1] T. Magdziarz,[2] J. Rathman,[2,3] J. Firman,[1] A. Mostrag,[2] V. Ribeiro,[2] B. Hobocienski[2]

[1]Liverpool John Moores University, Liverpool, United Kingdom; [2]MN-AM, Nürnberg, Germany and Columbus, OH, USA; [3]Ohio State University, Columbus OH USA

Motivation

- It is essential to identify substances that may cause adverse effects to the liver following systemic exposure.
- There are many mechanisms of action responsible for chemically-induced liver injury following exposure to xenobiotics.
- Many mechanisms of action for liver injury have been formalised into Adverse Outcome Pathways (AOPs).
- *In silico* toxicology can make use of knowledge of the Molecular Initiating Events (MIEs) in AOPs to identify potentially hazardous molecules and assist in grouping and read-across.
- Previously, Firman et al [1] established molecular sub-structural fragments associated with cholestasis following investigation of the MIEs in the AOP, see below

Alerts for Cholestasis (Firman et al. [1])

Oestrogenic steroid	**Sulfonamide (antimicrobial)**	**Azole antifungal**
Androgenic steroid	**Thiazide**	**Fluoroquinolone**
Stilbene derivative	**Benzenesulfonylurea**	**NSAID (-profen)**
Phenothiazine	**Beta-lactam**	**ACE inhibitor (peptidic)**
Dibenzocycloheptane	**Desosamine**	**Statin**

- In order to support prediction of organ-level toxicity following chronic exposure, structural knowledge needs to be transformed into usable and verified *in silico* tools.

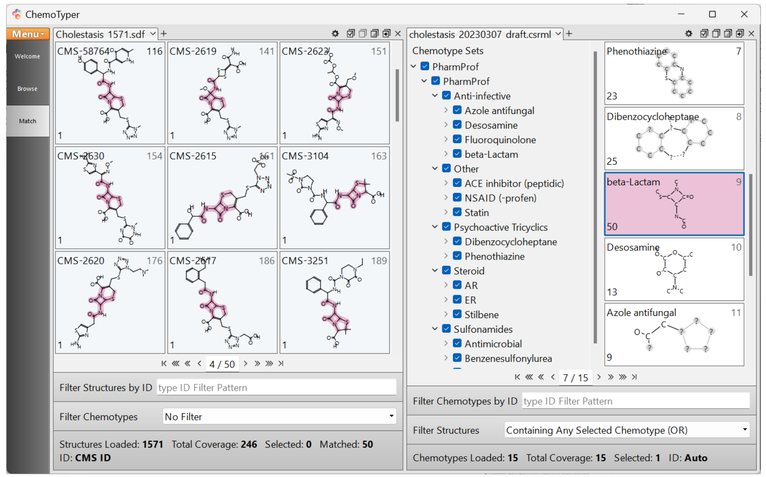
Aims

- The aim of this study was to develop a robust structure-based profiler for the identification of molecules potentially liable to induce cholestasis.
- Firman et al's [1] existing rulebase was extended through the use of an enriched data resource, allowing for the evaluation of the rules.

Methods

- Existing structural rules were coded in Chemical Subgraphs and Reactions Markup Language (CSRML) prior to inclusion in the ChemTunes software.
- A database of more than 1,500 cholestasis test results was curated and quality controlled.
- The cholestasis data were used to evaluate the chemotypes and calculate likelihood estimations to determine the strength of the predictability behind each alert.

Structure Group-Based Classification of Rules



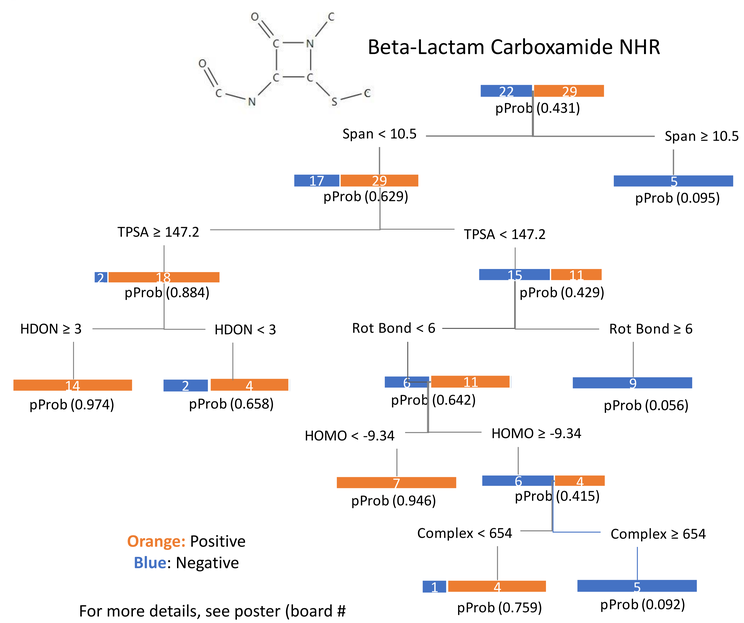
Machine Learning (ML) – Hybrid Rules [2]

- Optimize expert rules to improve reliability and extendibility, with the aim to:
  - ◊ Decrease the number of false positives
    - Many expert rules captured in fragments tend to give false positives in real use cases.
  - ◊ Increase the number of true negatives
    - Design rules to represent negative observations
- ML-Hybrid Rules for Cholestasis
  - ◊ Beta-lactam class included only 57% positives (pProb=0.43)
    - A mixture of positives and negatives.
  - ◊ A machine learning procedure can be applied to find the best combination of physico/molecular properties as well as other fragment features.
    - A recursive partitioning tree was prepared with selection and termination rules.
    - pProb indicates the conditional (Bayesian) probability of a rule having the structure fragment (beta-lactam) and properties.

Fragment Rules Statistics

Chemotypes	Z_score*	mean**	# hits
CHOL_Sulfonamides_Thiazide	**7.24**	**0.89**	19
CHOL_Psychoactive_Tricyclics_Azepines	6.67	0.76	25
CHOL_Anti-infective_beta-Lactam	6.23	0.57	51
CHOL_Anti-infective_Desosamine	**6.23**	**0.92**	13
CHOL_Psychoactive_Tricyclics_Phenothiazine	5.64	0.70	23
CHOL_Anti-microbial_Benzenesulfonylurea	5.22	0.83	12
CHOL_Steroid_Estrogenic	3.81	0.67	12
CHOL_Steroid_Estrogenic	3.81	0.67	12
CHOL_Other_ACE_inhibitor_peptidic	3.74	0.70	10
CHOL_Other_HMG_COA_reductase_Statin	3.11	0.58	12
CHOL_Steroid_Stilbene_derivative	2.69	0.67	6
CHOL_Steroid_Androgenic	1.70	0.42	12
CHOL_Anti-infective_Azole_antifungal	1.68	0.44	9
CHOL_Other_NSAID_-profen	1.12	0.33	15
CHOL_Anti-infective_Fluoroquinolone	0.20	0.24	17
CHOL_Antimicrobial_Sulfonamide	0.02	0.22	23
TXP_bond:CN_amine_pri-NH2_alkyl	-3.00	0.11	133
TXP_bond:COH_alcohol_pri-alkyl	-3.42	0.10	139

*Z-scores are adjusted by the size of each class against the full dataset.
**Mean of each class represents the sensitivity.
+ High mean and high Z-scores give confidence for a robust rule. Two structural chemotypes (thiazides and phenothiazines) are considered robust enough to be applied in real testing.
++Primary alkyl amines and primary alkyl alcohols are the root nodes that were further constrained to become negative cholestasis rules during the machine learning process. (TXP: ToxPrint Chemotypes)



For more details, see poster (board # P186 & Abstract #3701) on Tuesday.

Rules Overview

- The structural alerts for cholestasis were successfully coded into chemotypes using CSRML
- Chemotype rules for cholestasis have been classified according to pharmacological actions (see Figure and Screenshot)
- Chemotypes were refined further to define hybrid rules for cholestasis, taking molecular properties into account. Hybrid rules after machine learning process are represented in CSRML syntax.
- Performance statistics were calculated against the *in vivo* data.

Summary

- Alerts for cholestasis have been implemented in CSRML to create a value *in silico* tools for hazard identification and chemical grouping
- Alerts are optimized by the combination of physico-chemical properties or other structural fragments
- The rulebase provides valuable mechanistic information as an *in silico* New Approach Methodology (NAM)
- ML-Hybrid rule process strengthens the predictivity.

Next Steps

- Improve reliability and accuracy in validation process by reviewing the quality of the knowledgebase
- Classify the rules by mechanistic pathways beyond the structural confines
- Apply ML-Hybrid approach [2] more extensively (adopting a variety of ML techniques)
- Combine with NOEL/LOEL data to develop potency-aware rules

References

[1] Firman JW et al. (2021) Chem. *Res. Toxicol*. 34: 641–655
[2] Yang C et al. (2023) *Chem. Res. Toxicol*. In press

Acknowledgement

- Funding through the Cosmetics Europe Long-Range Science Strategy is gratefully acknowledged.