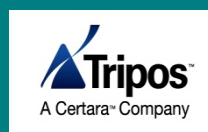
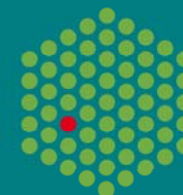


Mining Public Databases for Structure-Activity-Relationships

Bernd Wendt
Ulrike Uhrig
Fabian Boes



EMBL

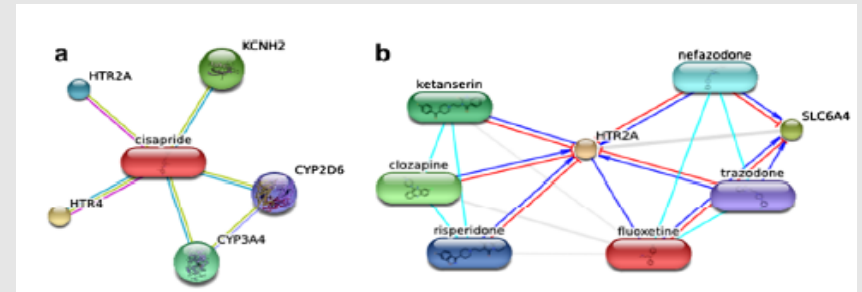


Problem to solve

- Target-based drug discovery
 - Focus on specific protein (classes, e.g. kinases)
 - Quick optimization on-target
 - High attrition in later stages (off-target effects)
 - Potential lack of relevance of target protein
- Phenotype-driven drug discovery
 - Focus on specific cell state transitions (e.g. apoptosis)
 - Optimization without bias regarding specific targets
 - Target identification problem: MoA unknown (on-targets)
- Strategies for target identification needed
 - Identification of on- and off-targets

On- and Off-Target Predictions

- “Bridging Chemical and Biological Space” aka “Pharmacoinformatics” or “Chemogenomics”
 - Protein-centric
 - Map spaces on basis of similarities among proteins
 - Ligand-centric
 - Map spaces on basis of similarities among drugs
- Several approaches available, e.g.
 - STITCH (Bork-Group, EMBL):
 - Network of binary interactions
- **Need more than binary interactions to derive Structure-Activity-Relationships (SARs)**

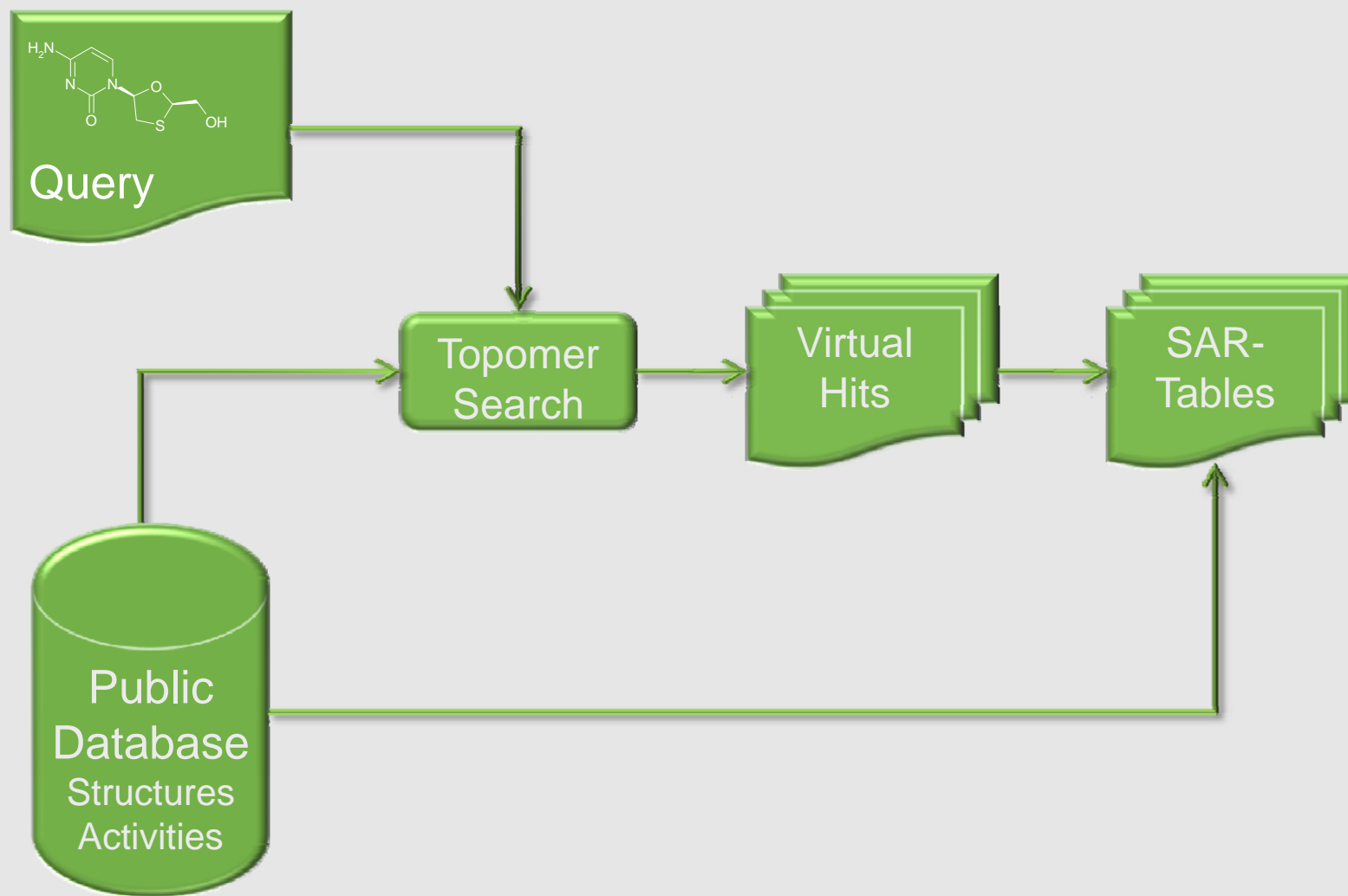


Kuhn, M. et al. Large-scale prediction of drug-target relationships FEBS Lett. 282, 2008, 1283-1290

Outline

- Predicting on- and off-target effects
 - Use structures and activities from public domain
 - Identify structure-activity-relationships (SARs) using automated 3D-QSAR (Topomer CoMFA)
 - Apply SAR-models to propose changes for enhancing or attenuating the off-target effect
- Generalization of the approach

Workflow: Modeling off-target effects



Challenge: How to identify Structure-Activity Relationships ?

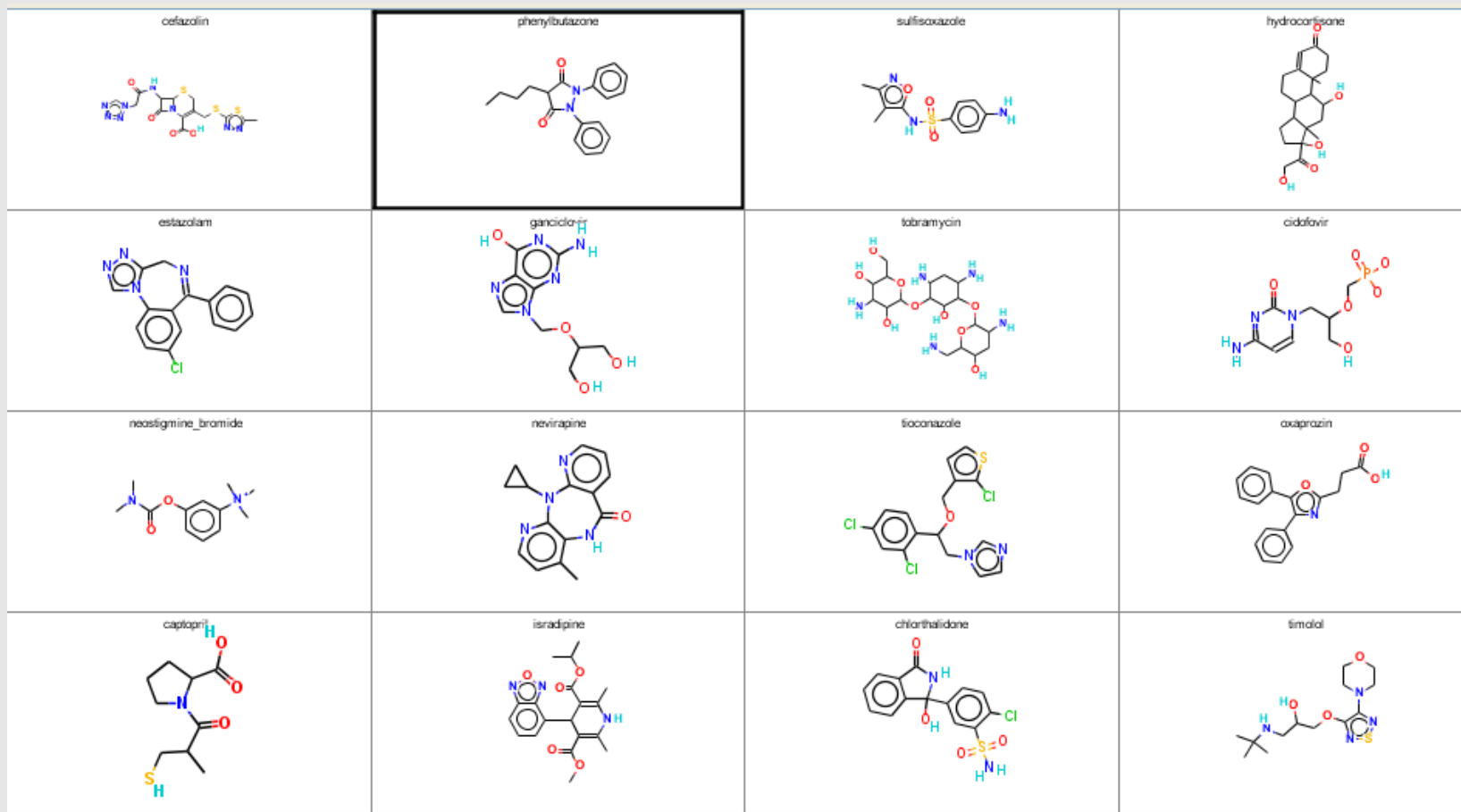
- Can a reliable SAR-model be constructed from any SAR table?
 - Or a subset thereof?
- Use QSEA-method: Quantitative Series Enrichment Analysis
- Outcome-1: SAR-model with maximum structural coverage and good statistical performance
- Outcome-2: Prediction of test compounds

Wendt, B.; Cramer RD., Quantitative Series Enrichment Analysis (QSEA), J. Comp. Aid. Mol. Des. 2008, 22, 541-555

Generalization of the approach

- Query
 - Selection of query structures
- Candidates
 - Gather structures and activities from public databases
- Search
 - Virtually screen public databases using queries
- Assemble SAR-tables
- Identify Trends
 - Systematic analysis (QSEA) of SAR-table
- Presentation of selected examples

Structural Queries: Jain-set of 255 structures covering 22 different biological targets



This dataset can be downloaded <http://www.jainlab.org/downloads.html>

Cleves, A.E.; Jain, A.N.: J. Med. Chem, 2006, Vol.49, 2921-2938 (Robust Ligand-Based Modeling of the Biological Targets of Known Drugs)

Candidates: Contents of Public Databases

Database	Structures (associated with data)	Assays/ Targets	Comment
PubChem	39.323.334	1273 (assays)	
ChemBank	8.795.367	1744 (assays)	
Binding-DB	46.723	618 (enzyme names)	Mixed binding data (K _i , IC ₅₀ , ..) per target and mixed target source organisms
K _i -Database	27.743	546 (enzyme names)	Activity spread per a single structure too big. Different literature sources may not be comparable
Binding MOAD	3.410	426 (EC numbers)	Target data uniformity. E.C. too coarse grained for SAR
Affinity-DB	748	474 (PDB ids)	Data uniformity

All databases were accessed December 2008

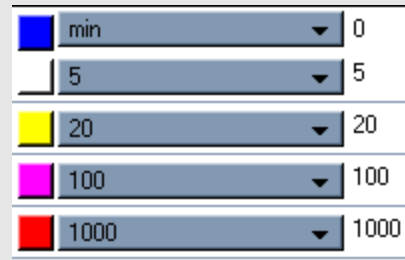
<http://pdsp.med.unc.edu>
<http://www.bindingdb.org>

<http://www.bindingmoad.org>
<http://chembank.broad.harvard.edu>

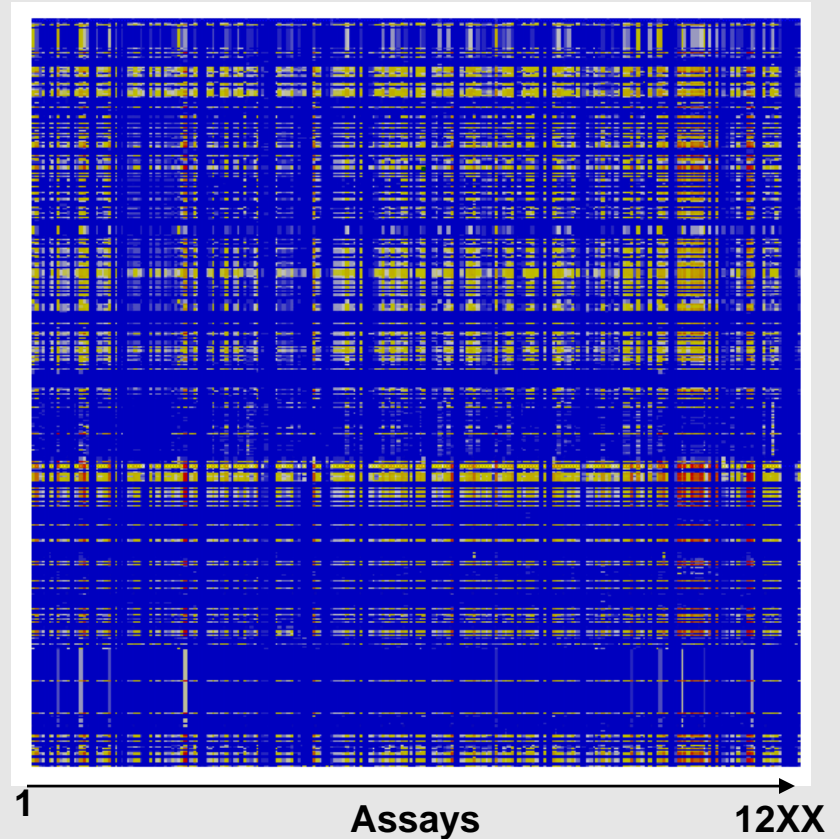
<http://pc1664.pharmazie.uni-marburg.de/affinity/>
<http://pubchem.ncbi.nlm.nih.gov>

PubChem: Virtual Screening Results

- Heatmap shows number of hits with biological data per query and assay
- Decent number of hits in Pubchem for many queries

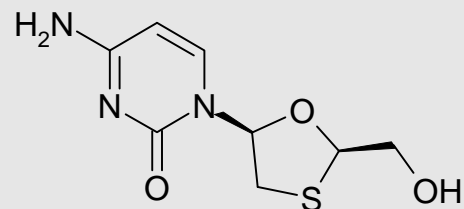


queries



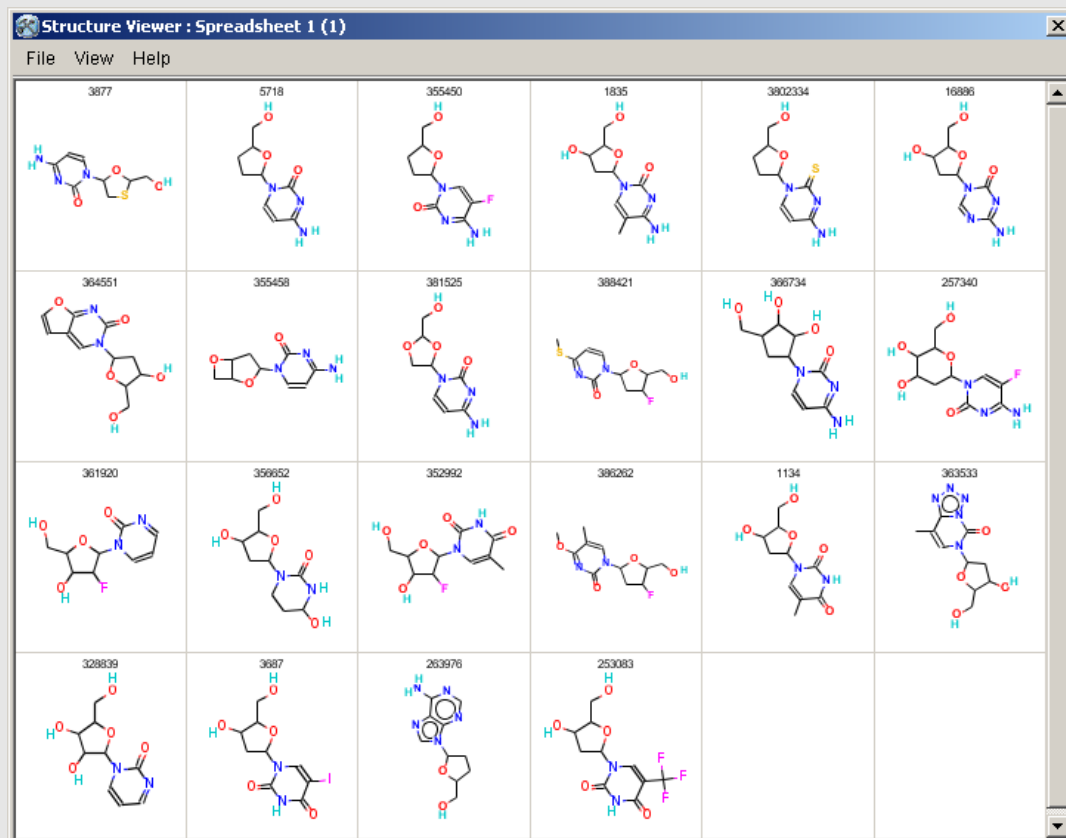
Assemble SAR-table: Example

- Lamivudine

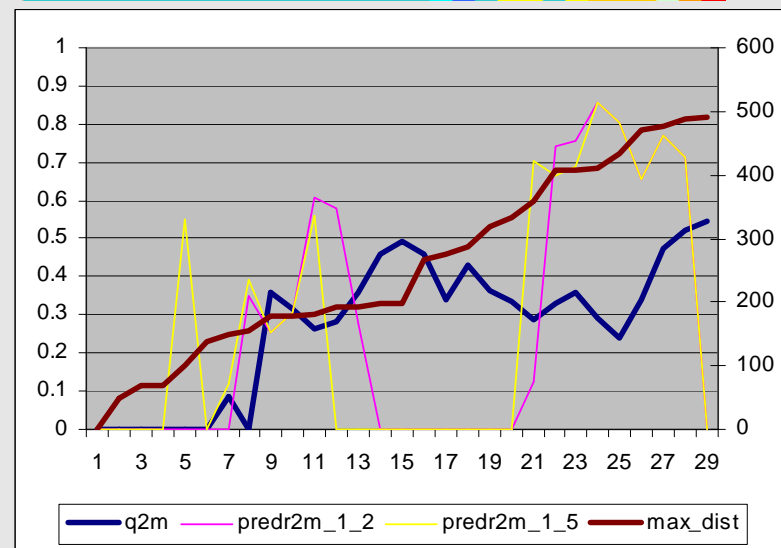
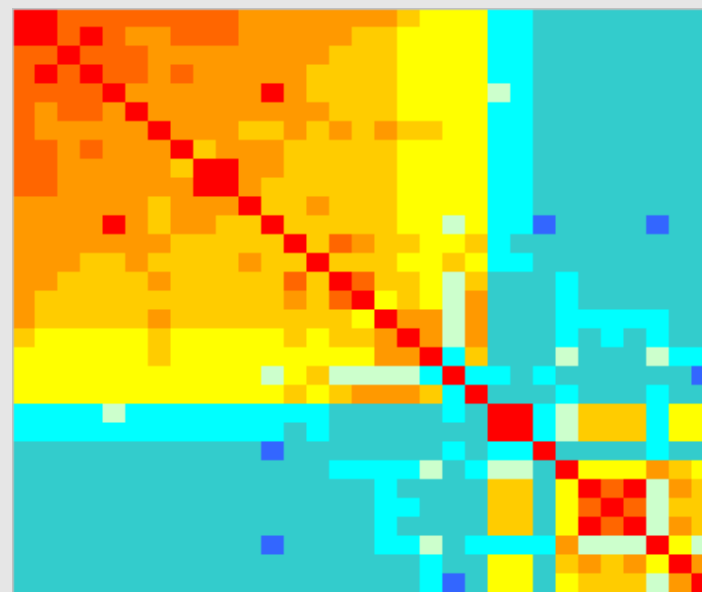
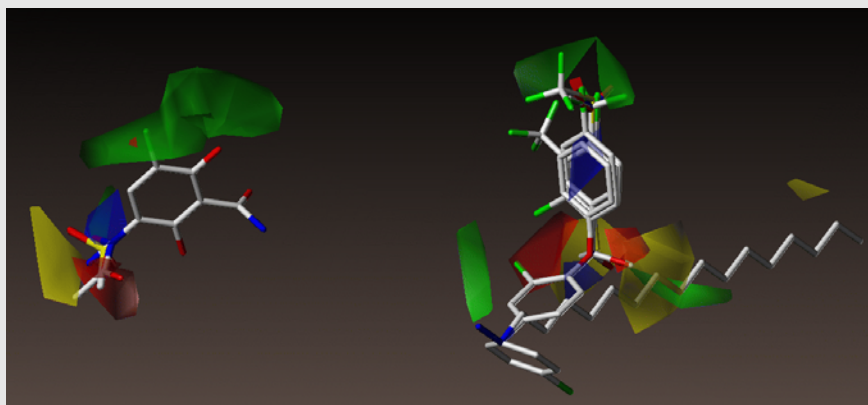
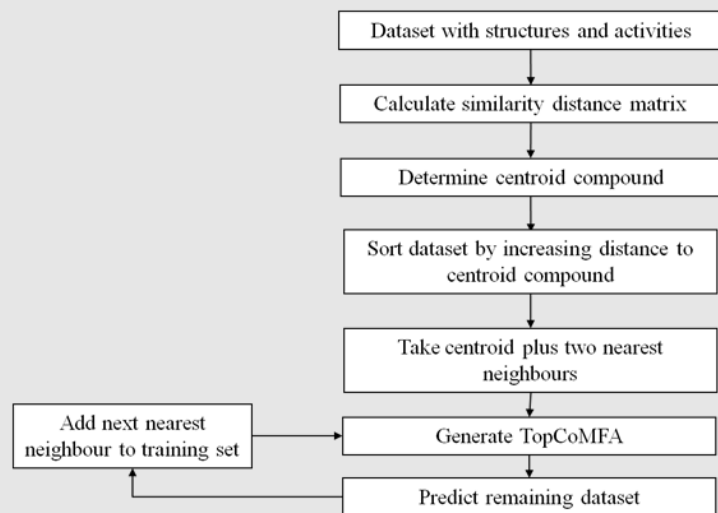


Lamivudine

- potent nucleoside analog reverse transcriptase inhibitor (nRTI)
 - approved by the FDA in 1995
- SAR-table:
 - 22 hits with associated biological data
 - 1 exact match
 - NCI cancer screen
 - Activity reported in $-\log(GI50)$

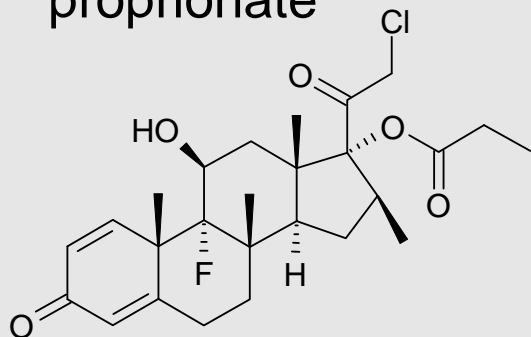


QSEA



Results: Example I: CYP 3A4 Inhibition

- Clobetazol proprionate

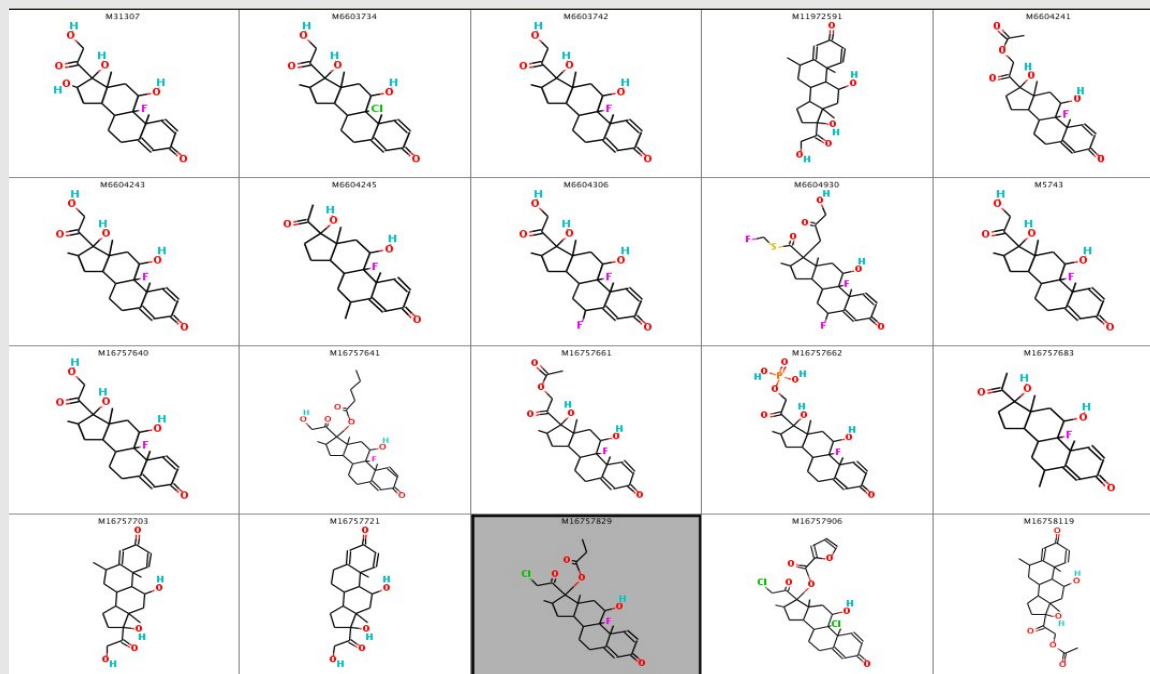


Clobetazol proprionate

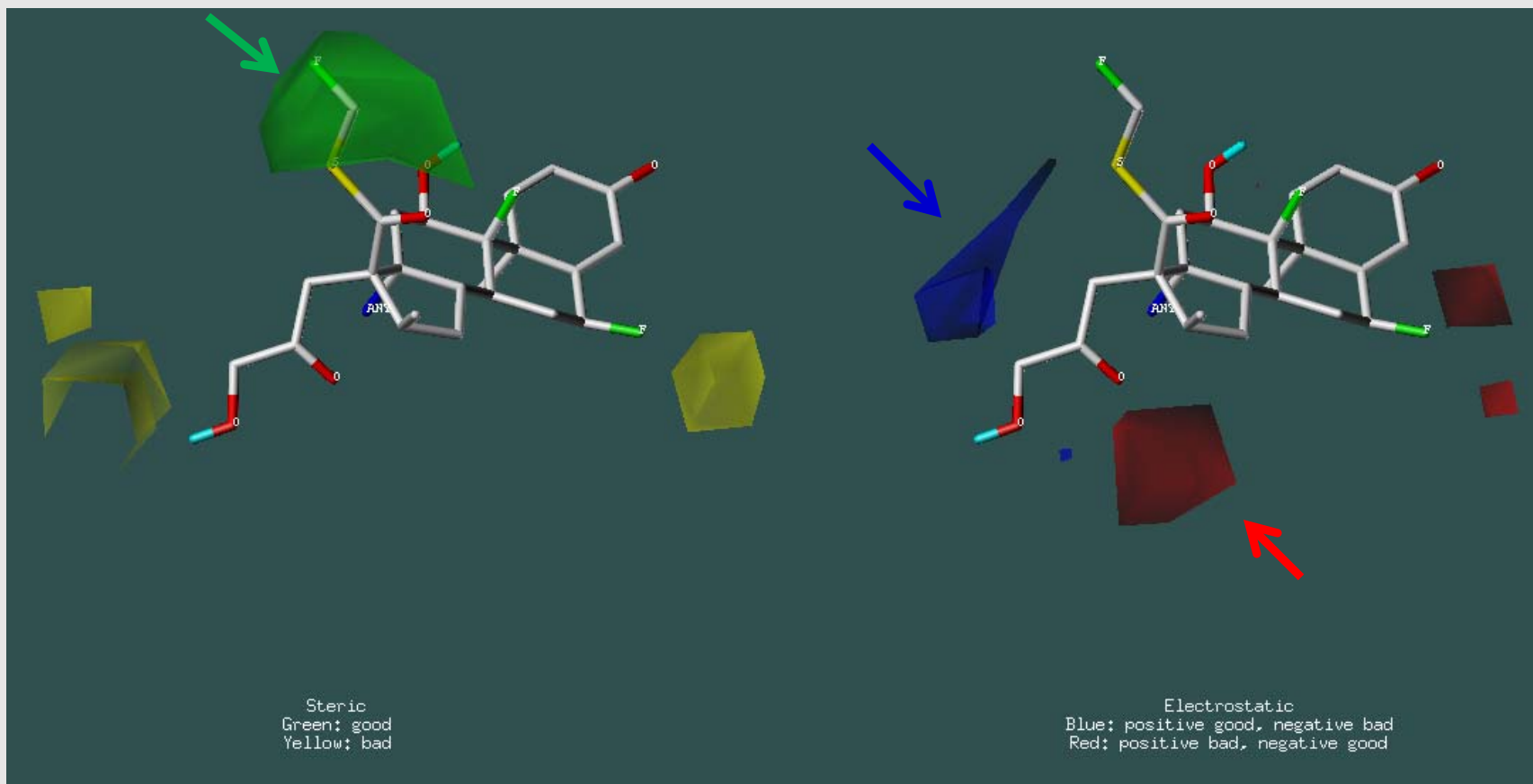
- Corticosteroid used to treat various skin disorders

SAR-table:

- 30 hits with associated biological data
- 1 exact match
- CYP 3A4 screen
- CoMFA model with 20 neighbours

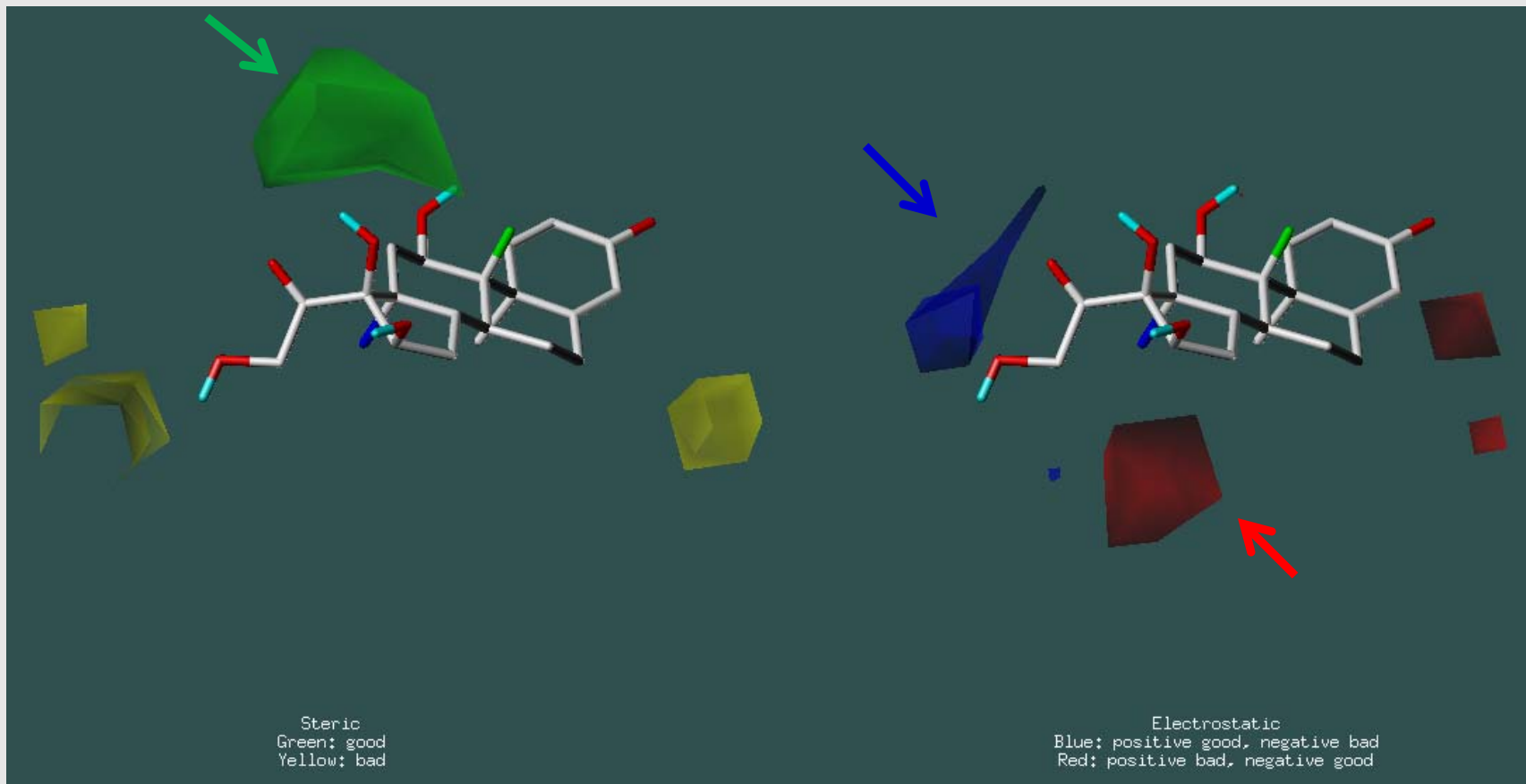


Results: Example I: CYP 3A4 Inhibition



Active!

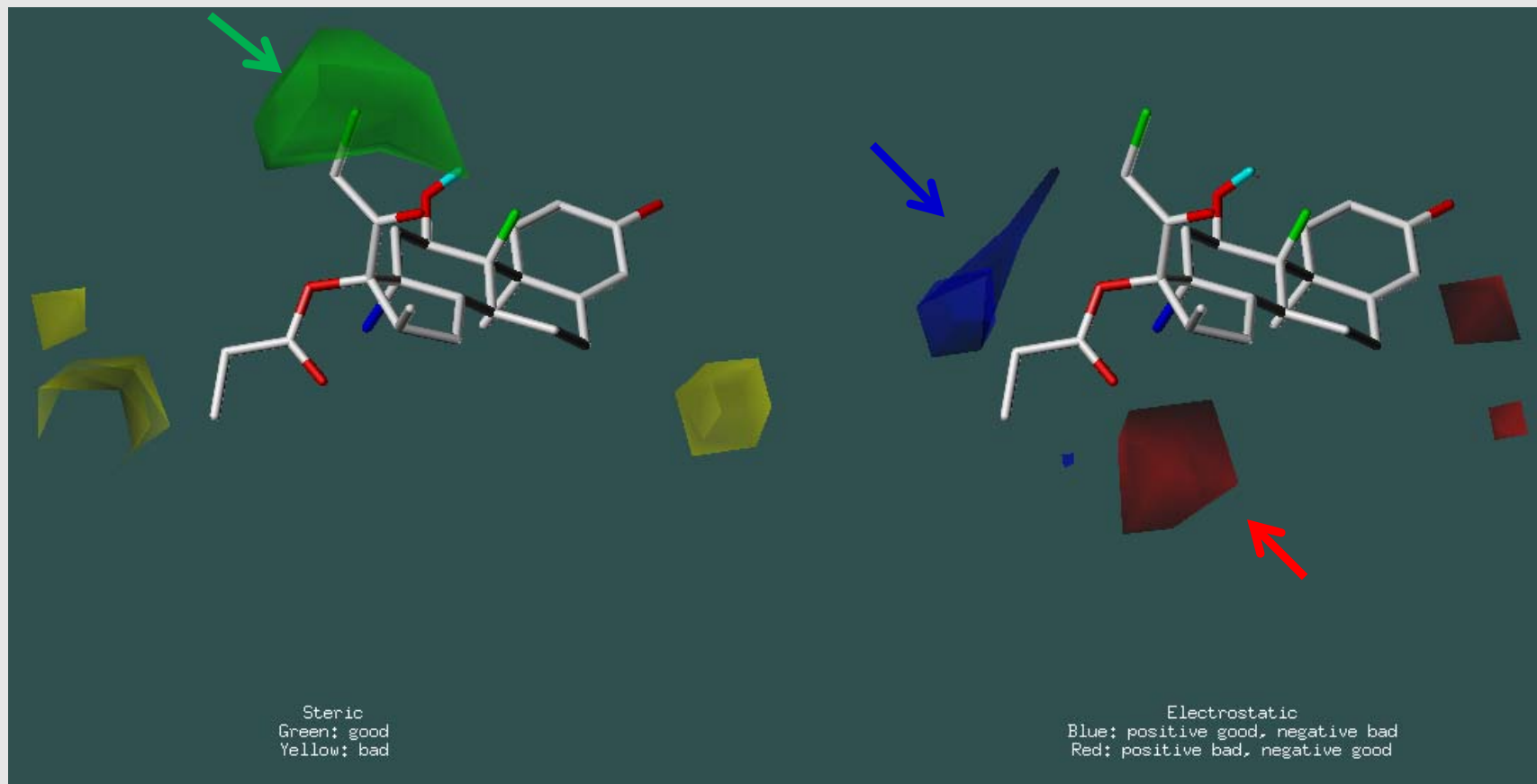
Results: Example I: CYP 3A4 Inhibition



Inactive!

Results: Example I: CYP 3A4 Inhibition

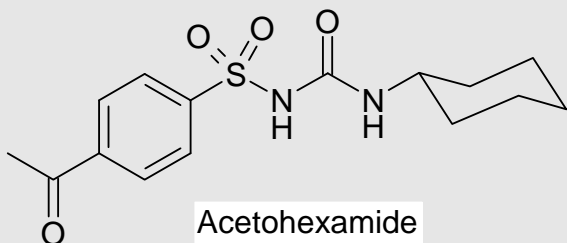
Is Clobetasol propionate active ?



Yes, it is active ?

Results: Example II: CYP 2C19 Inhibition

- Acetohexamide

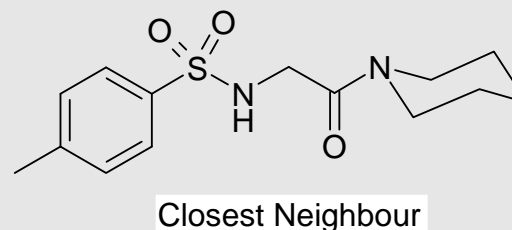


- sulfonylurea medication used to treat diabetes mellitus type 2

- SAR-table:

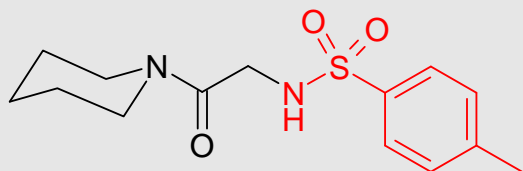
- 39 hits with associated biological data
 - no exact match
- CYP 2C19 screen

- Closest Neighbour found:



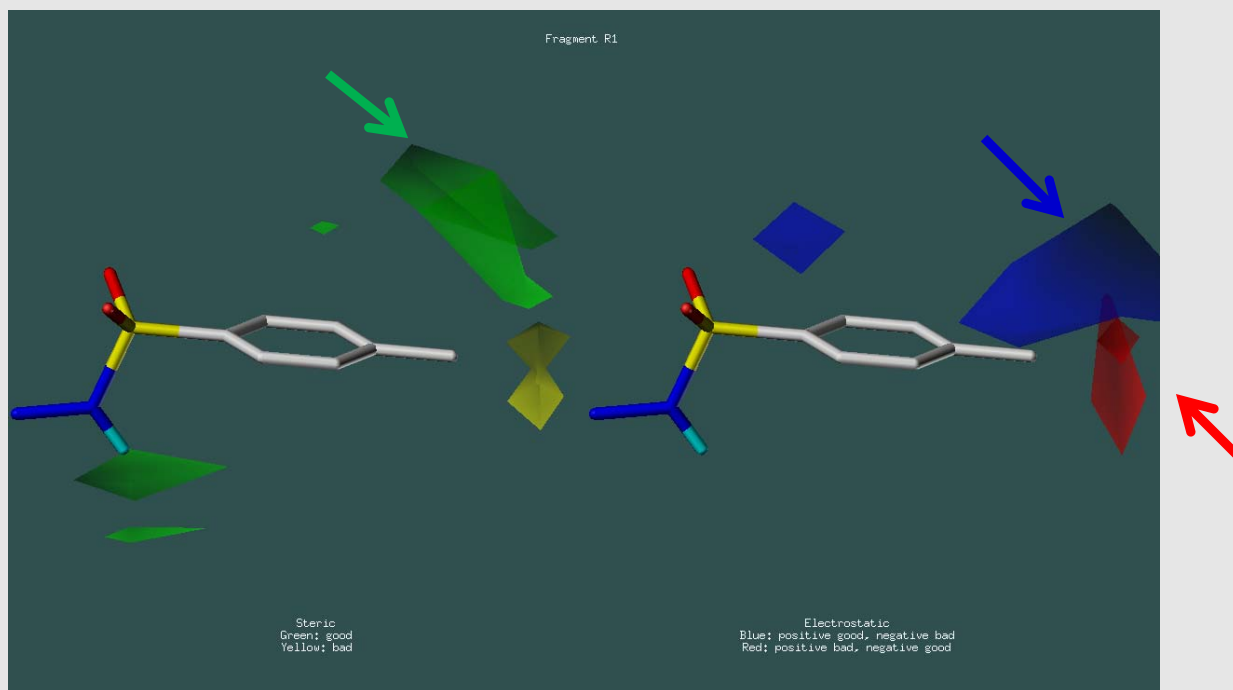
- No activity in
- CYP 2C19 screen
- Possible conclusion:
 - Acetohexamide is also inactive

Results: Example II: CYP 2C19 Inhibition



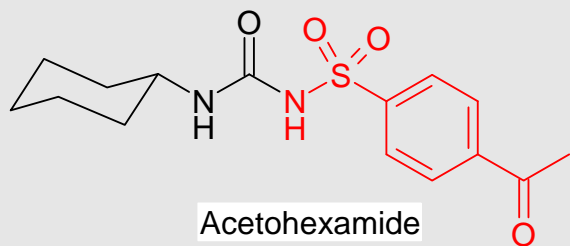
Closest Neighbour

Inactive

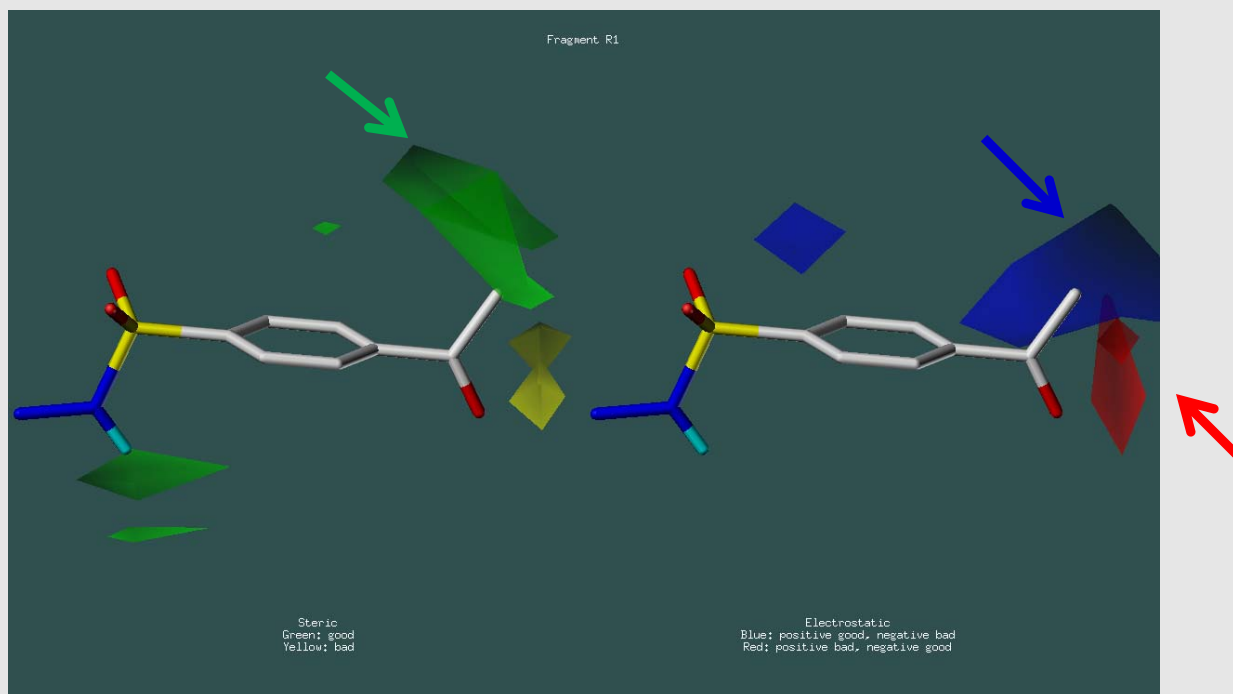


Left-hand side (R1)

Results: Example II: CYP 2C19 Inhibition

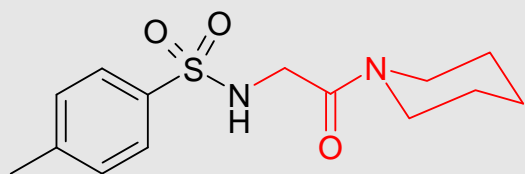


Active!!



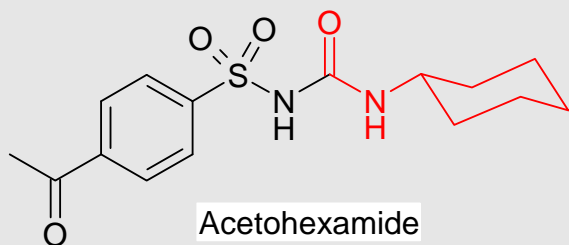
Left-hand side (R1)

Results: Example II: CYP 2C19 Inhibition



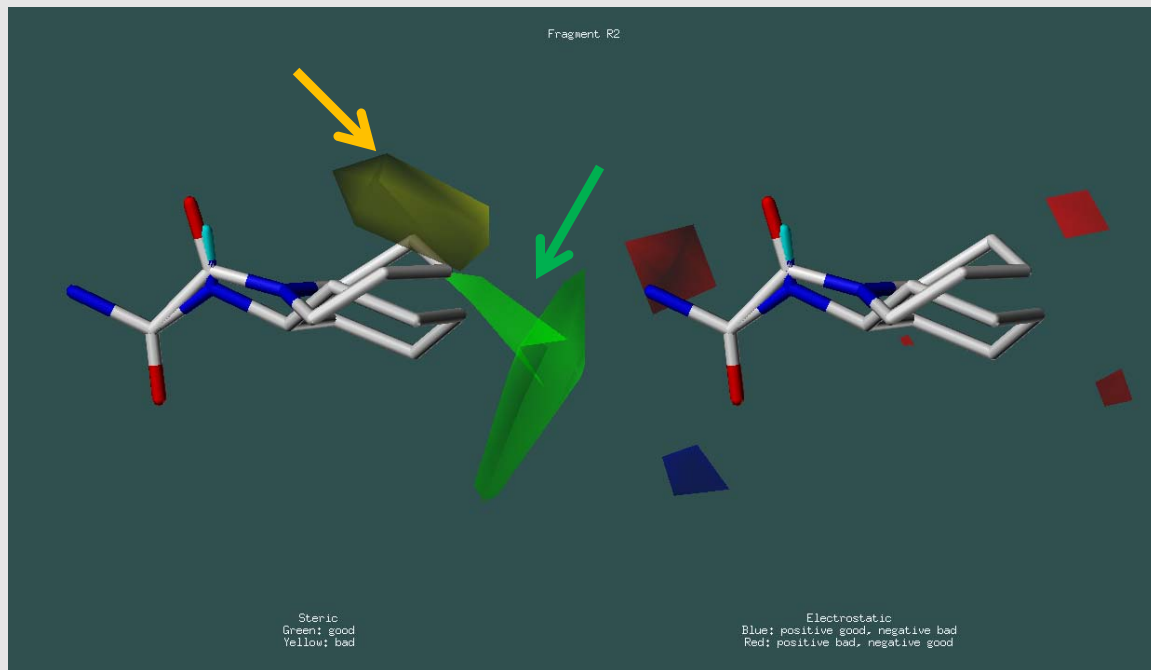
Closest Neighbour

Inactive



Acetohexamide

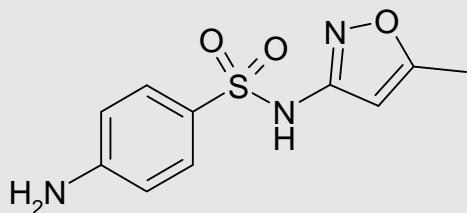
Active



Right-hand side (R2)

Results: Example III: Estrogen Receptor activation

- Sulfamethoxazole

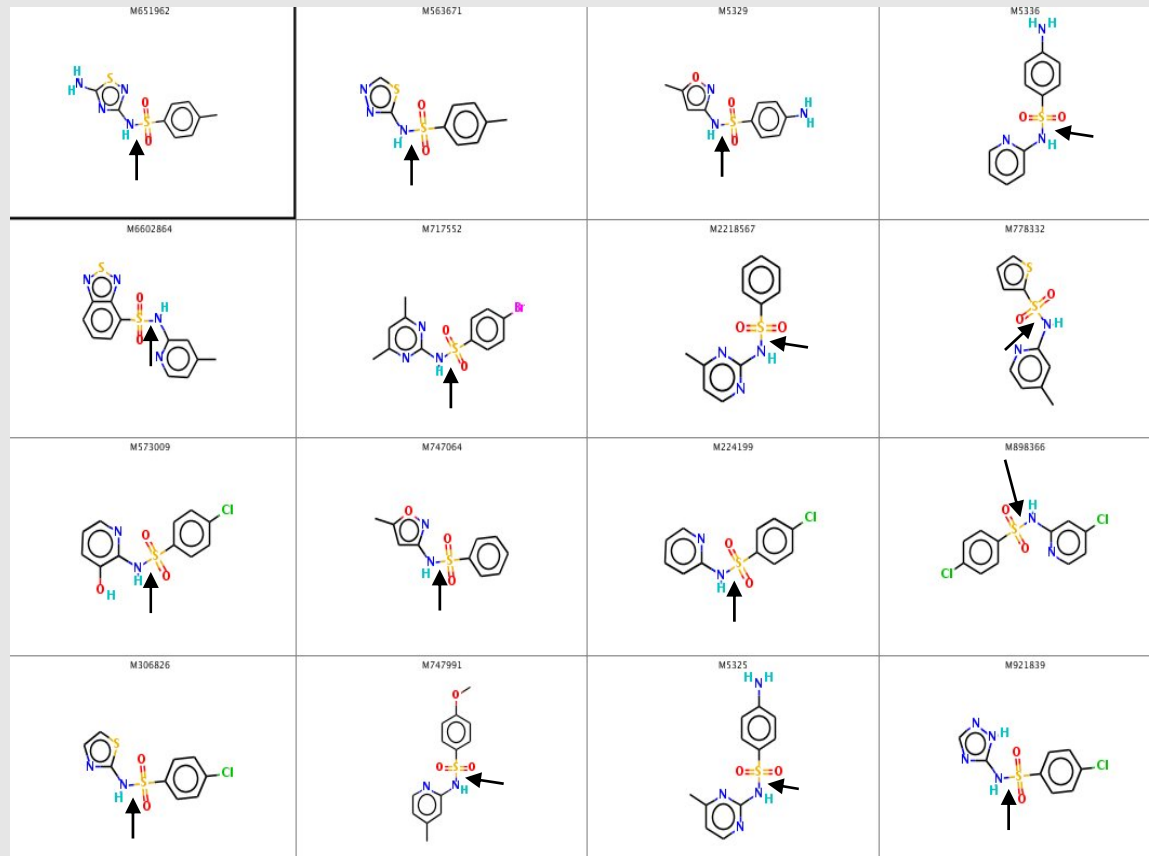


Sulfamethoxazole

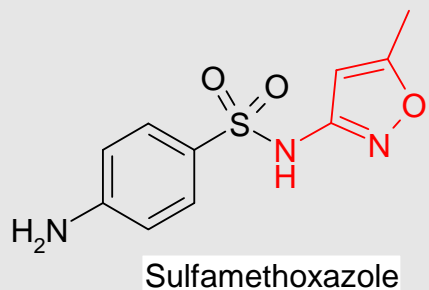
- Sulfonamide bacteriostatic antibiotic
- Antagonist of PABA
- Competitively inhibits dihydropteroate synthetase (tetrahydrofolate synthesis pathway)

- SAR-table:

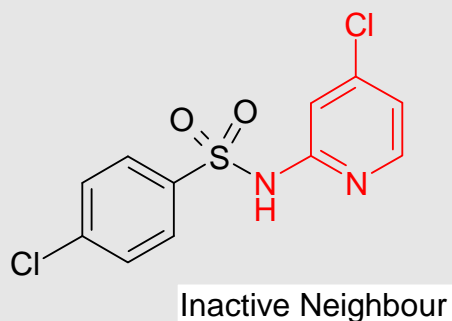
- 127 hits with associated biological data
 - 1 exact match
- ER alpha screen
- SAR-model with 16 close neighbours



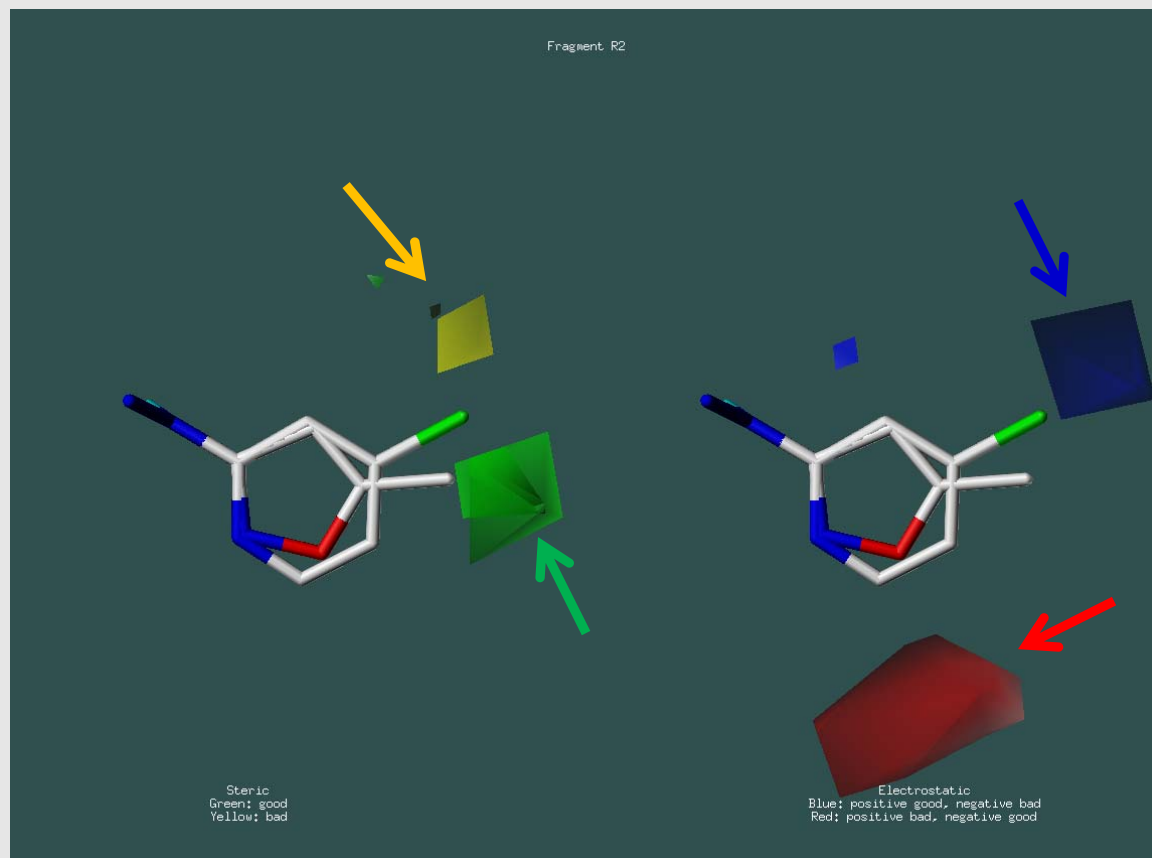
Results: Example III: Estrogen Receptor activation



Active

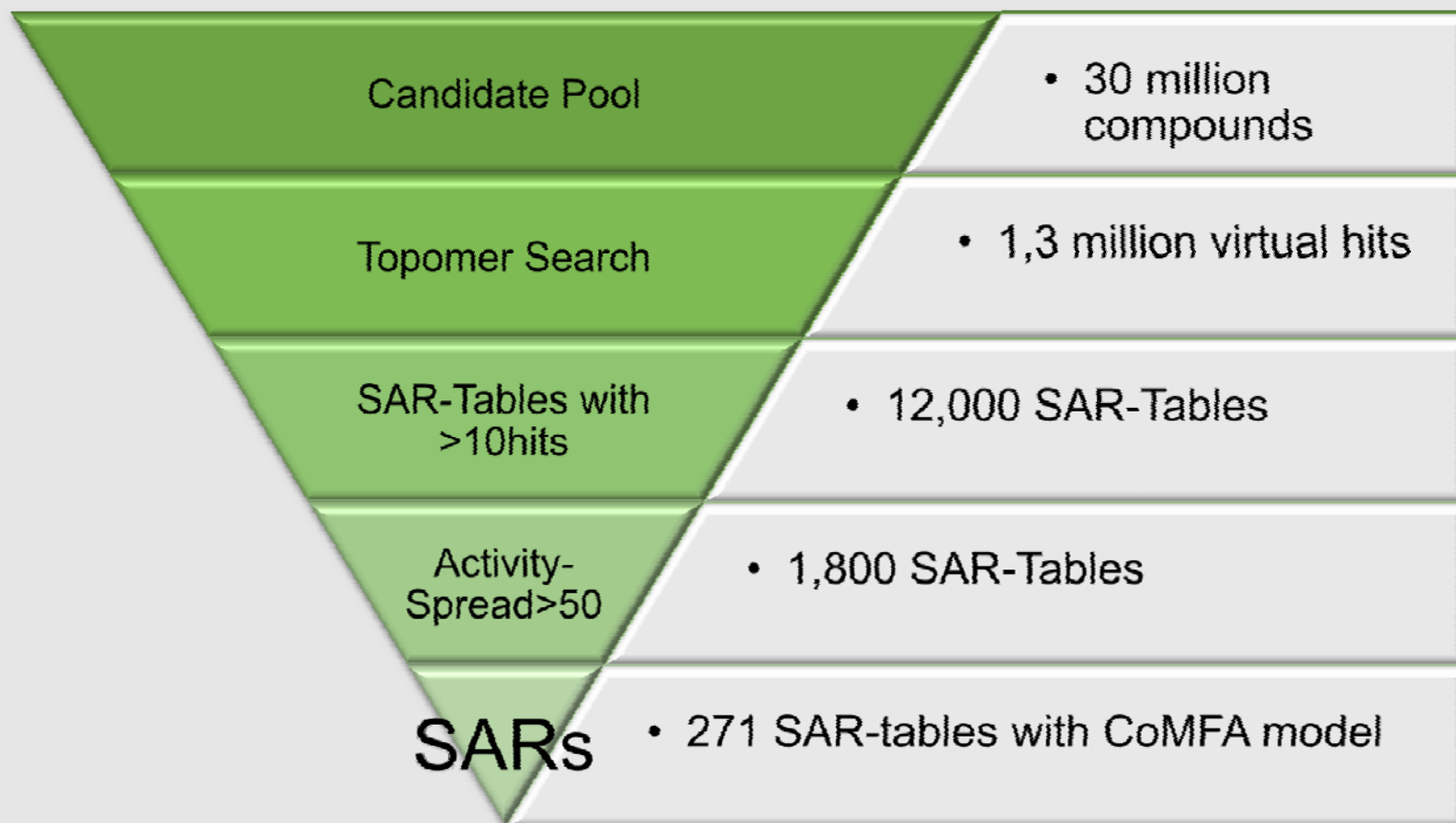


Inactive

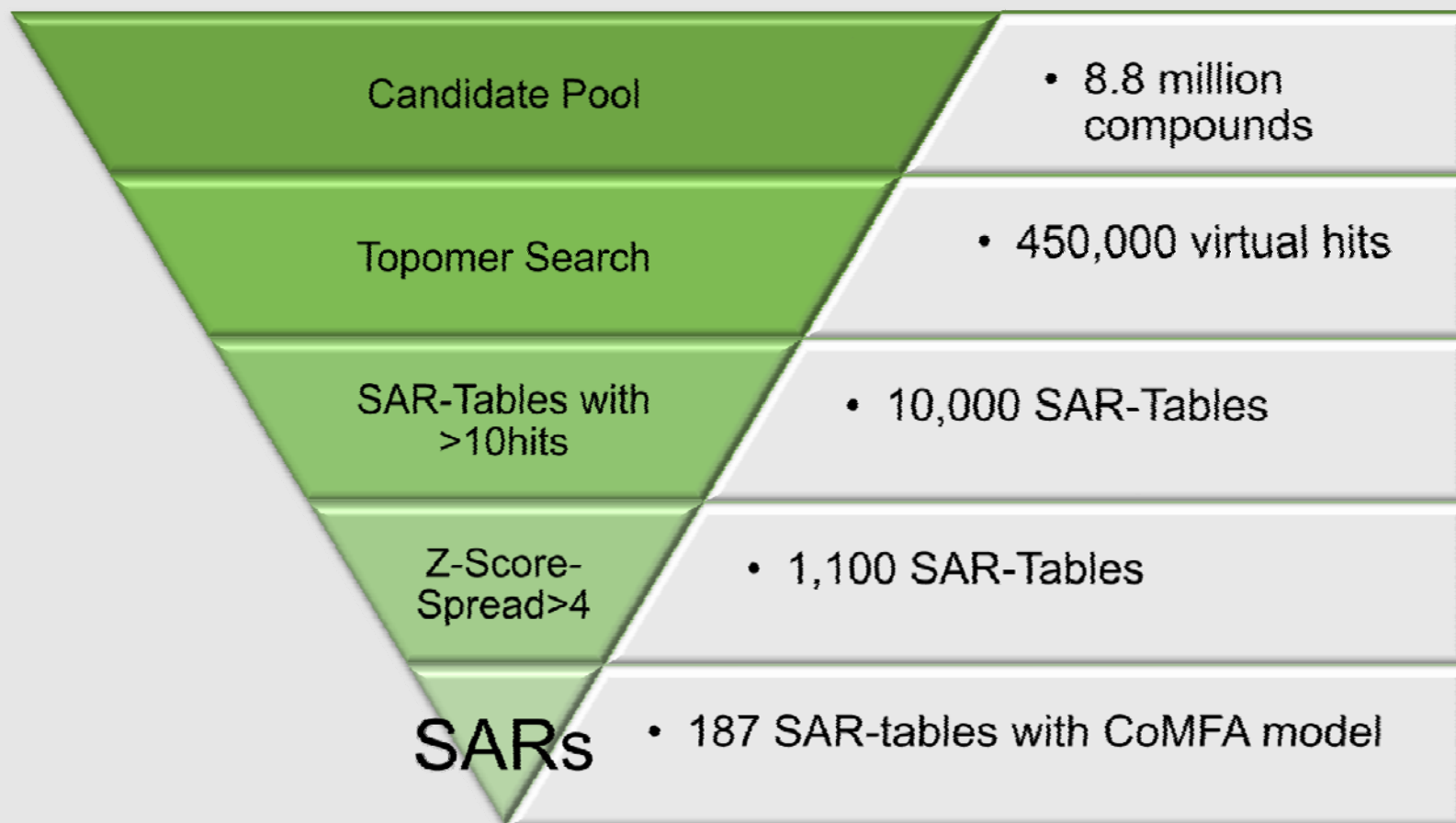


Right-hand side (R2)

Pubchem-Results: From Candidate Pool to SAR



ChemBank-Results: From Candidate Pool to SAR



Summary

- Representative set of marketed drugs used to mine two of the largest public databases for SAR-trends
 - Mining for SAR-Trends is possible
- PubChem and ChemBank useful sources for SAR information
- Would be interesting
 - to apply approach to curated databases or corporate databases
 - Use queries with more current chemistries
 - Higher hit rates expected
- Proposed as new tool for target identification

Acknowledgements

- EMBL/Elara
 - Joe Lewis
 - Christoph Schultes
 - Marcel Mülbaier
 - Bernd Janssen

 - *Jochen Ammenn*
- Tripos
 - Dick Cramer
 - Fred Soltanshahi
 - Brian Campbell
 - Brian Masek
- All people who were involved in setting up and maintaining:
 - PubChem
 - ChemBank