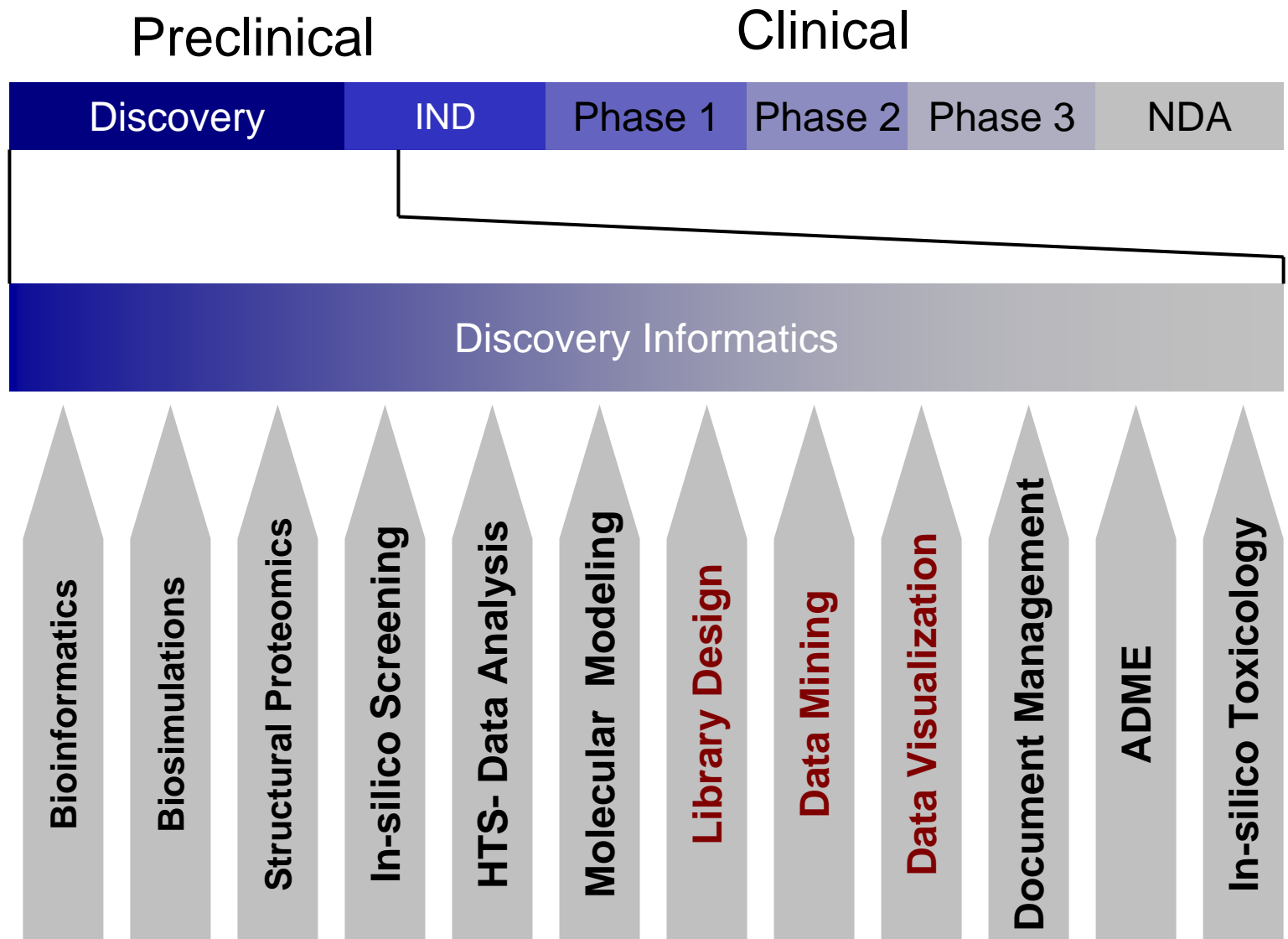


Substructural Analysis of Toxicological databases



Hugo O. Villar
Mark R. Hansen
Jason Hodges
Robin Friedman

Drug Discovery Cycle



Descriptor Based Informatics

Molecular Descriptor

(1,0,1,0,0,0,1,0,0,...)

3.25, 7.24,.....

Toxi/Pharmacophores

3D structural data

Biological data

Test System:AMES SALMONELLA
TYPHIMURIUM

Strain Indicator:TA100

Metabolic Activation:NONE

Method:PREINCUBATION

Dose:3.3-333 UG/PLATE (TEST MATERIAL
SOLVENT: ACETONE)

Results:NEGATIVE

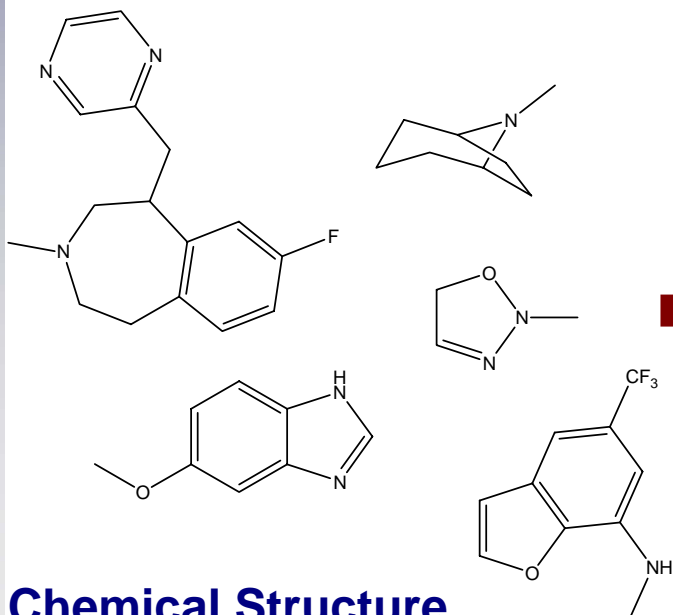
Reference:

Test System:AMES SALMONELLA
TYPHIMURIUM

Strain Indicator:TA100

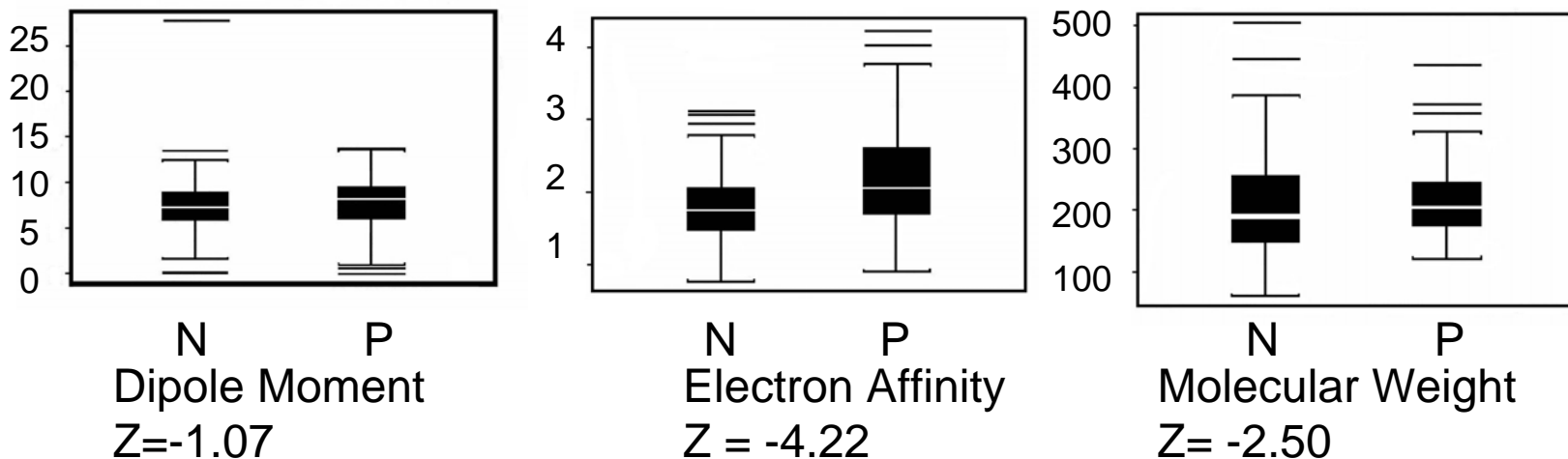
Metabolic Activation:HAMSTER, LIVER, S-9,
AROCLOL 1254 (10 OR 30%)

Method:PREINCUBATION



Chemical Structure

Property Based : Ames Test



Ames positive compounds have a tendency to have:

- larger number of hydrogen bond donors
- higher molecular weights
- higher electron affinities

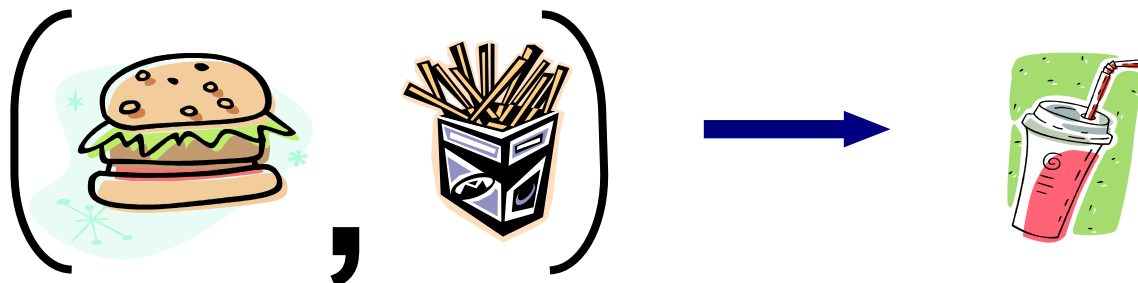
How do these results translate into better leads?

Deconvolution requires experts

Analysis in terms of substructures

Association Rules

- Look at relations between items
 - Bypass the use of properties
 - Primary information in the database
- Identify relations between sets of items



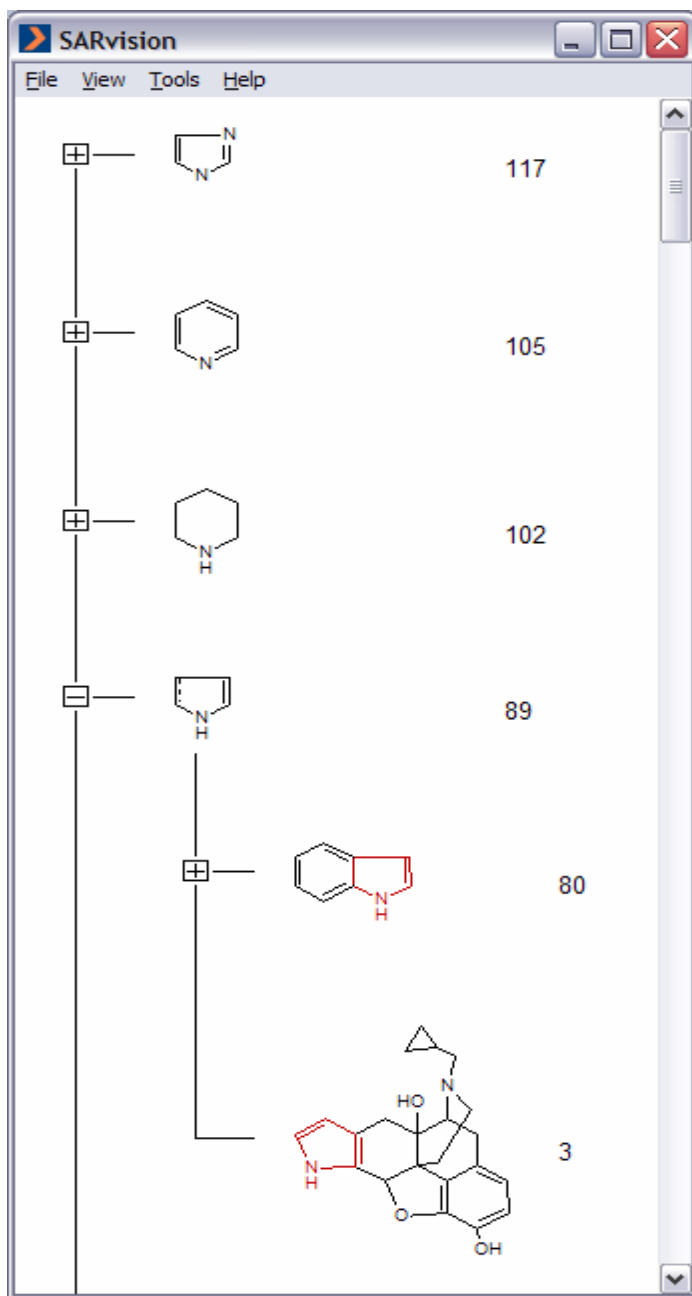
(Aromatic, Amine) => Ames Positive

Our itemsets are substructural elements

Substructures as Itemsets

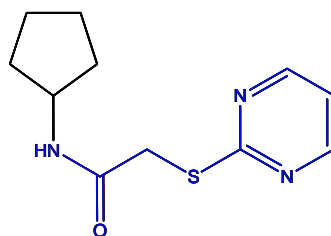
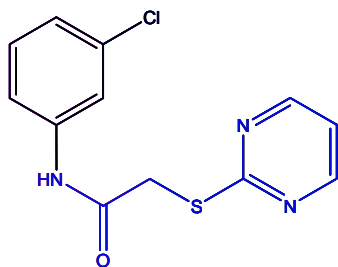
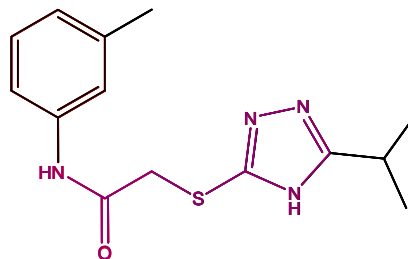
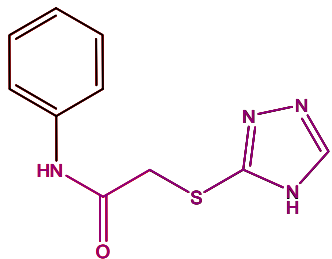
- Large number of substructures
 - Trp has 675 unique moieties*
- Hard to establish relations among them
 - Parent/child relations
 - Not all are relevant
- Raw lists are hard to manage
- Perception of Chemical Scaffolds
 - Establish lists
 - Use knowledge base

SARvision

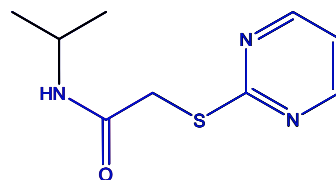
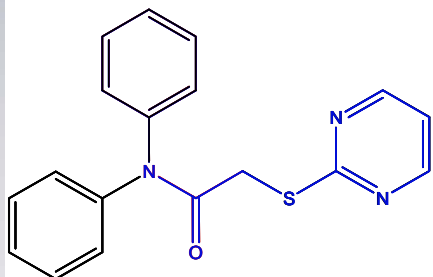


- Identify scaffolds
 - common to a set
 - non-trivial
 - chemically distinct
 - chemically meaningful
- Organize hierarchically
- Correlate with properties
- Carry out counts
 - occurrence
 - frequency in actives

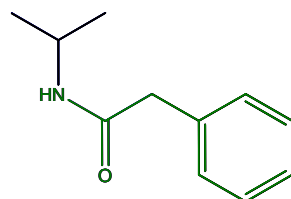
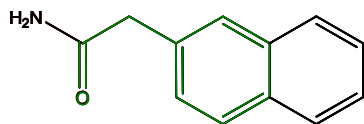
Knowledge Based Scaffold Perception



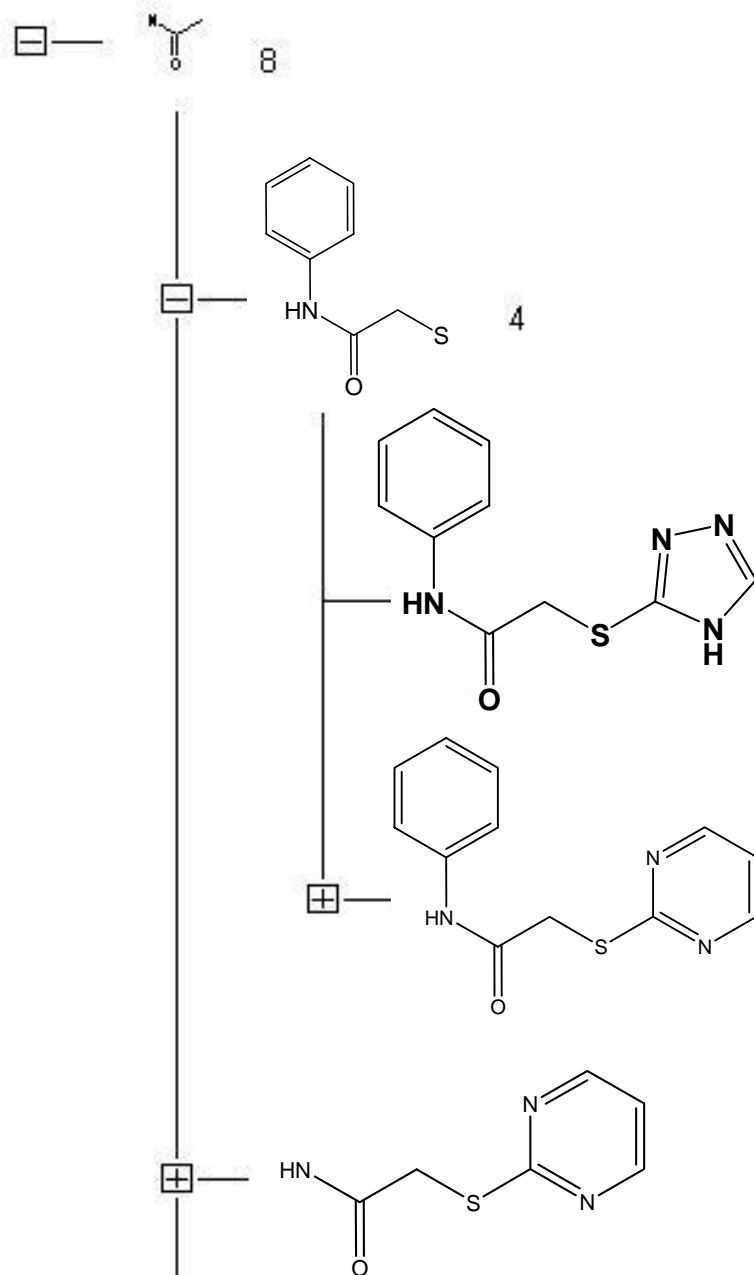
Identify scaffolds present



Increasing complexity



Non trivial extensions



Hierarchical Organization

Amide as Parent

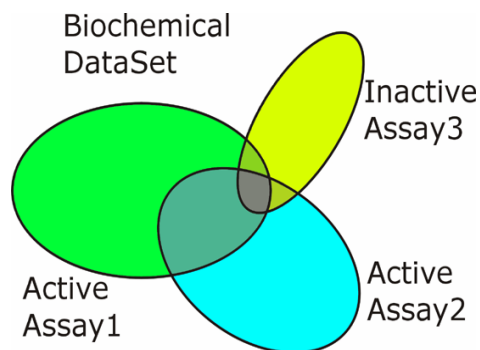
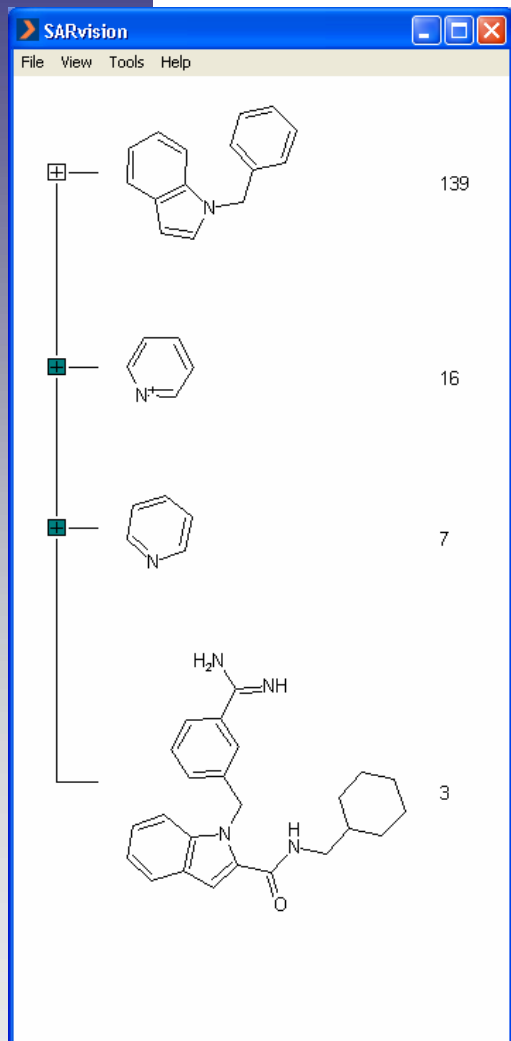
Non Trivial Extension

Recursively expanded

Frequency/ Counts

Proportion

Easily Coupled Chemical and Biological Data



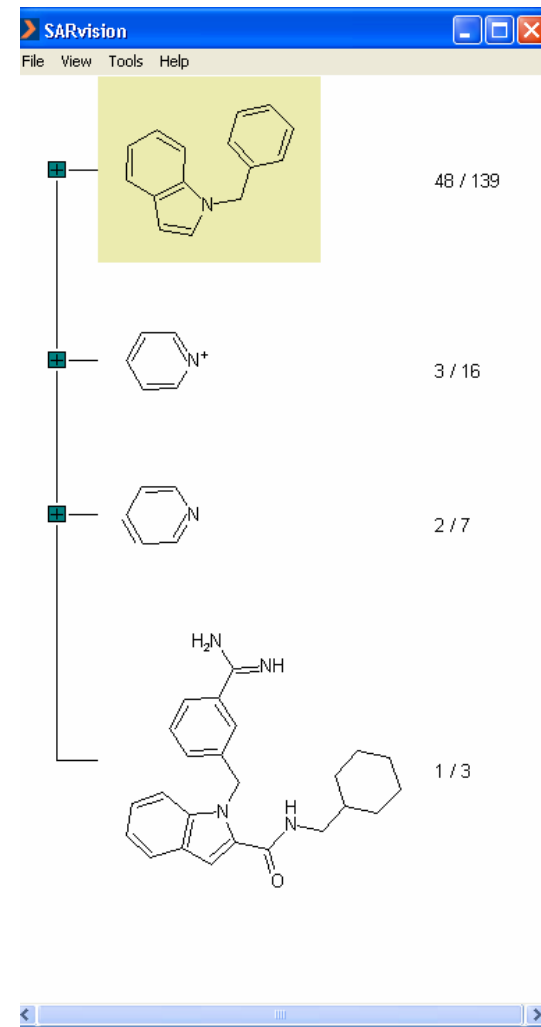
Select Active Properties

Add property: Add

Property name:	Lower bound:	Upper bound:
TA97_Activated_RESULTS	<input type="text" value="0"/>	<input type="text" value="1"/>
TA100_RESULTS	<input type="text" value="0"/>	<input type="text" value="1"/>
TA1537_RESULTS	<input type="text" value="0.0"/>	<input type="text" value="0.0"/>

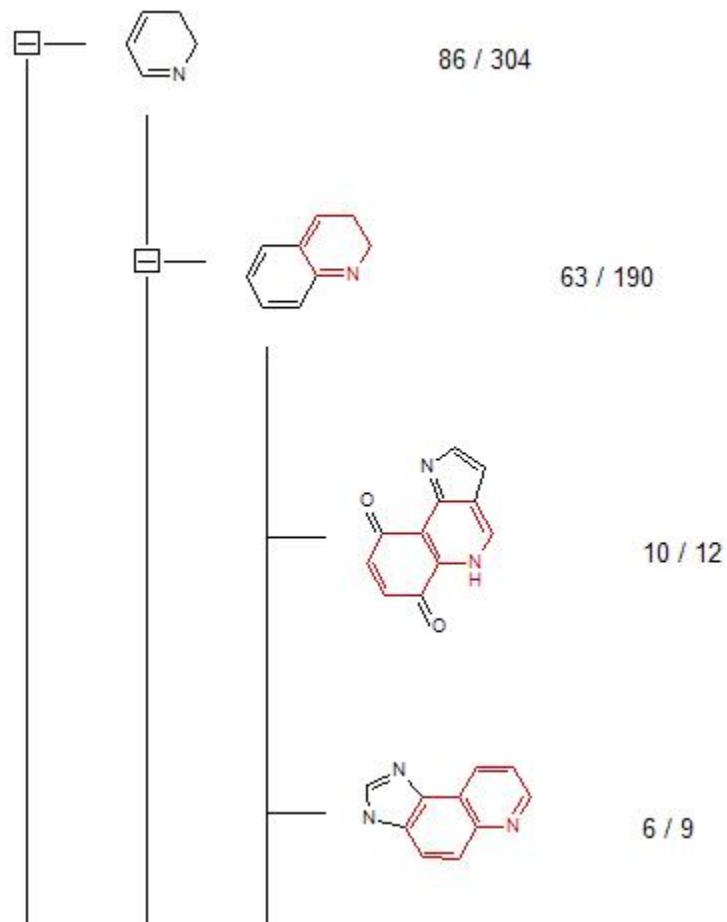
Properties selected: 3/14

OK Cancel Clear << Prev Next >>



Ames Data Sets

- **AMES test dataset.**
 - 6,039 molecules from CCRIS / TOXnet.
- Identify scaffolds prevalent in each class
- Determine the frequency
- Identify rules
 - Need to find support



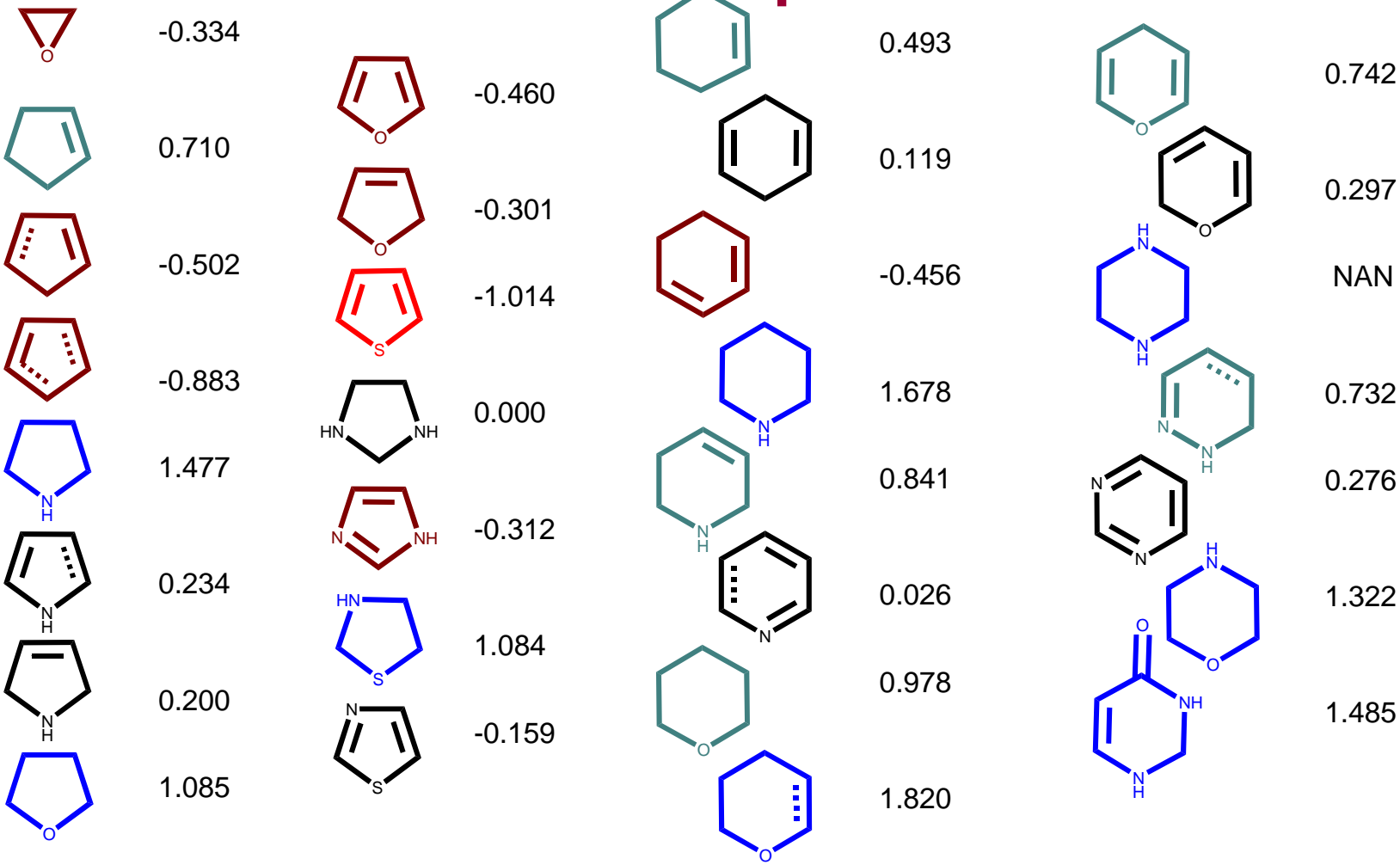
Determining Support for Rules

- Odds Ratio:

$$OR = \frac{a_{pos} / a_{neg}}{b_{pos} / b_{neg}}$$

- $\pm 1.96 \sqrt{1/a_{pos} + 1/a_{neg} + 1/b_{pos} + 1/b_{neg}}$
- Confidence Interval of Proportions
 $\pm 1.96 \sqrt{((1-P) P / n)}$
- Common in epidemiological/ clinical studies
- Not so common in chemoinformatics work

Ames test: Simple Scaffolds

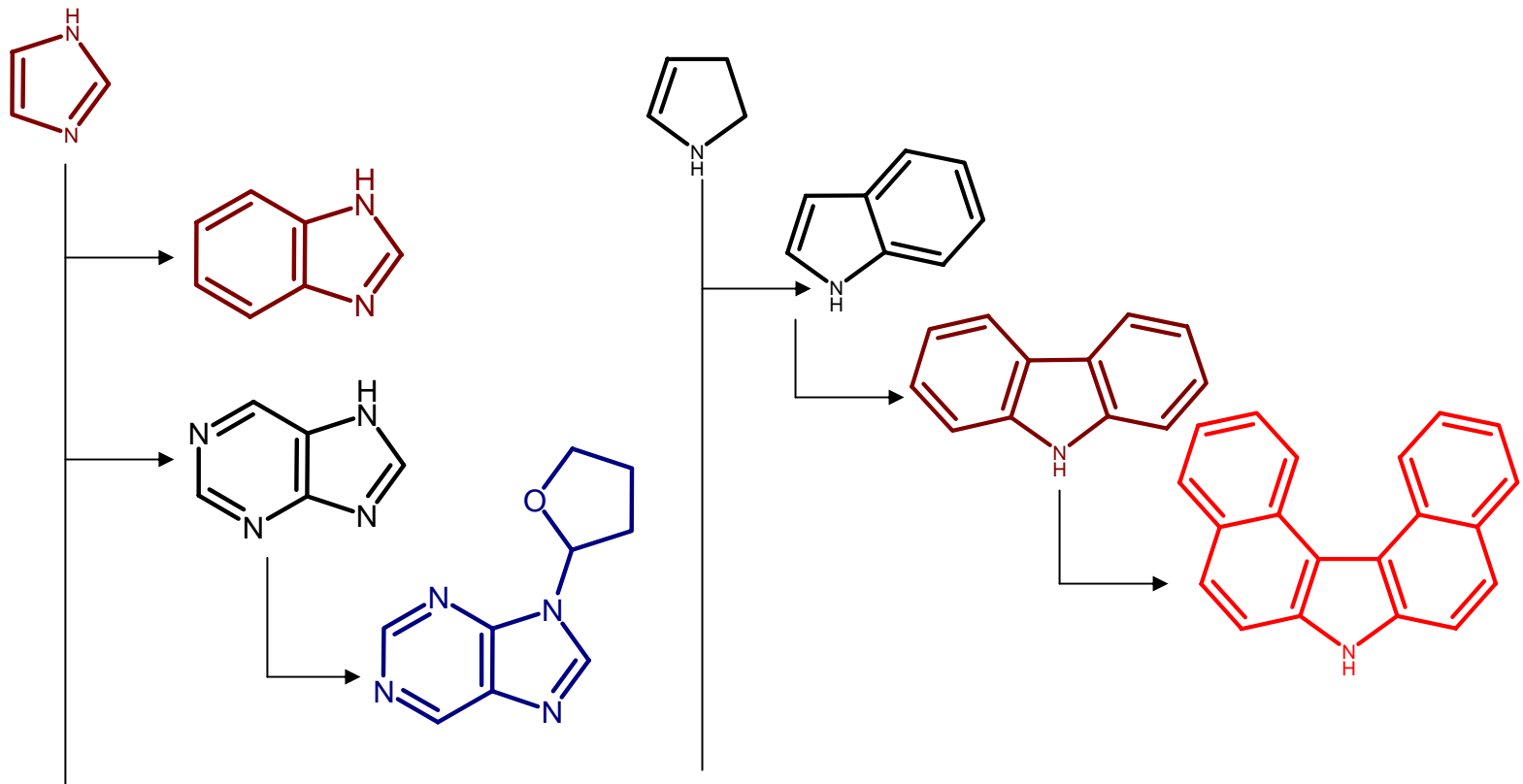


Log (OR)

■ >1.0;
 ■ (0.3, 1.0);
 ■ (-0.3, 0.3);
 ■ (0.3, 1.0);
 ■ <-1.0

Ames Test: Familiar Relations

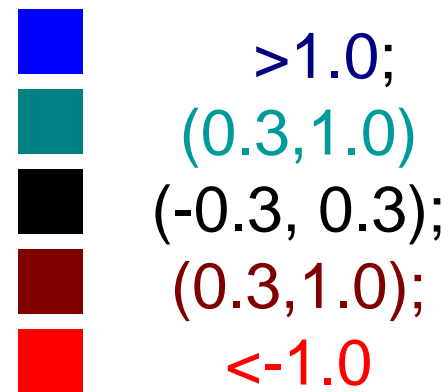
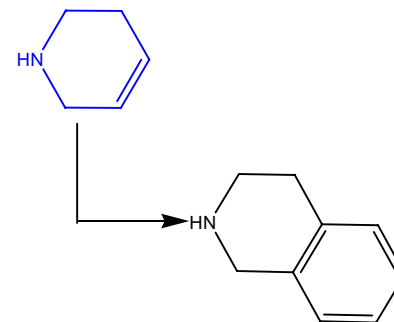
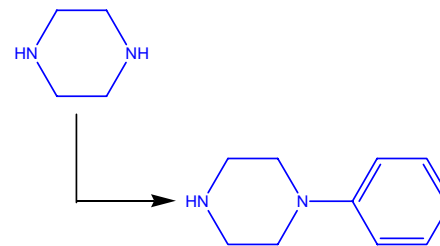
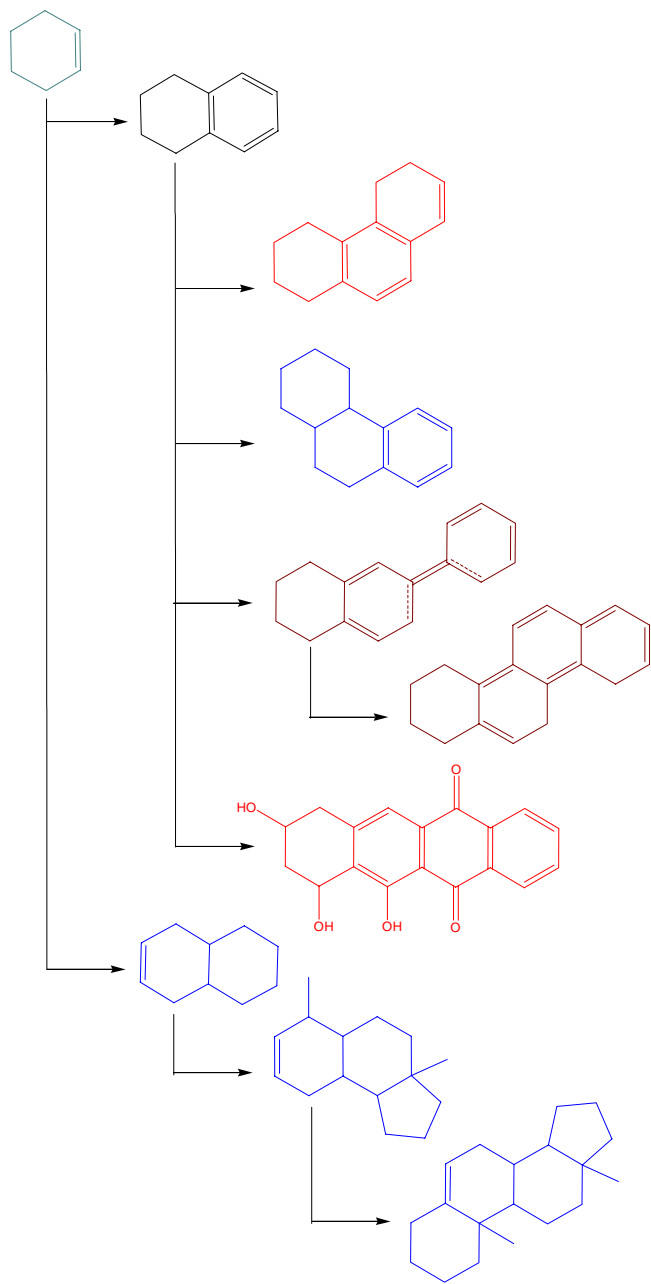
Build up of structures can alter the odds of being positive



Log (OR)



Ames Test: Familiar Relations



Relevance for Drug Design

3882 Drug



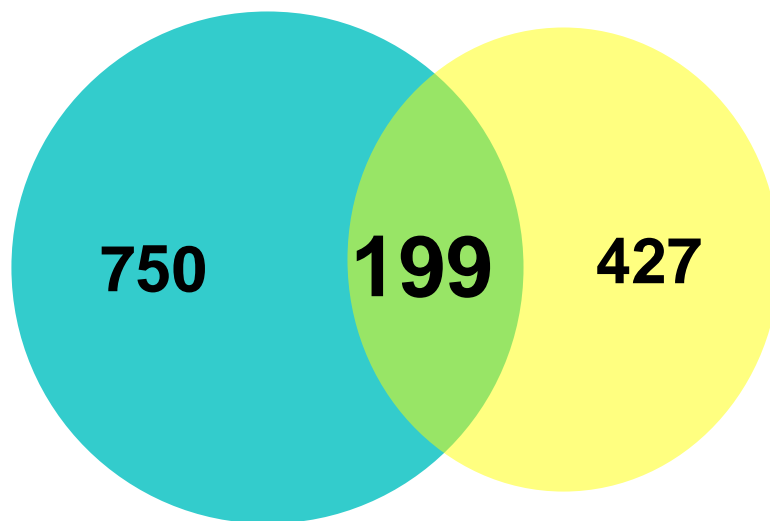
Drugs Scaffolds

6039 Ames Tested Compounds



Ames Scaffolds

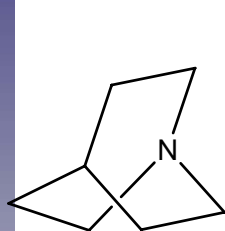
Common to at least 5 molecules



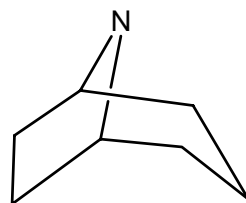
**Public domain datasets show significant redundancy
Datasets have marginal overlap with drug-like compounds**

Examples of Scaffolds

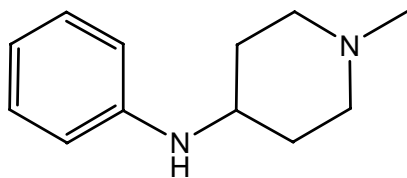
In drugs:



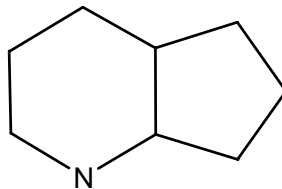
Pancopride,
Quinidine
Dolasentron



Cocaine
Scopolamine
Atropine



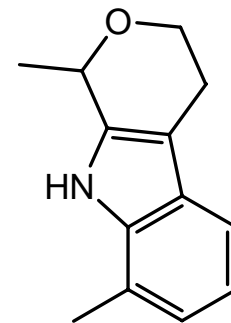
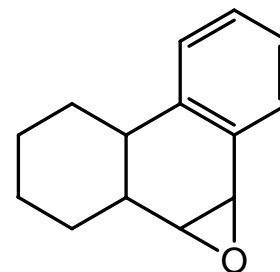
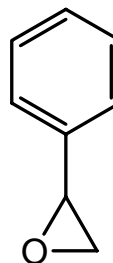
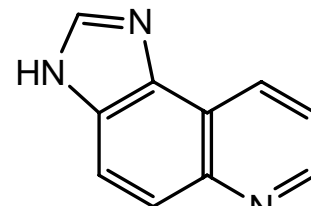
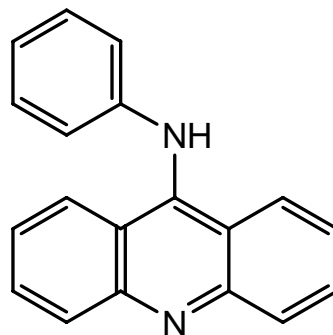
Lofentanyl
Lorcainide
Thenalidine



Rodocaine
Ajmaline
Estrazinol

not in Ames Databases

In AMES Database

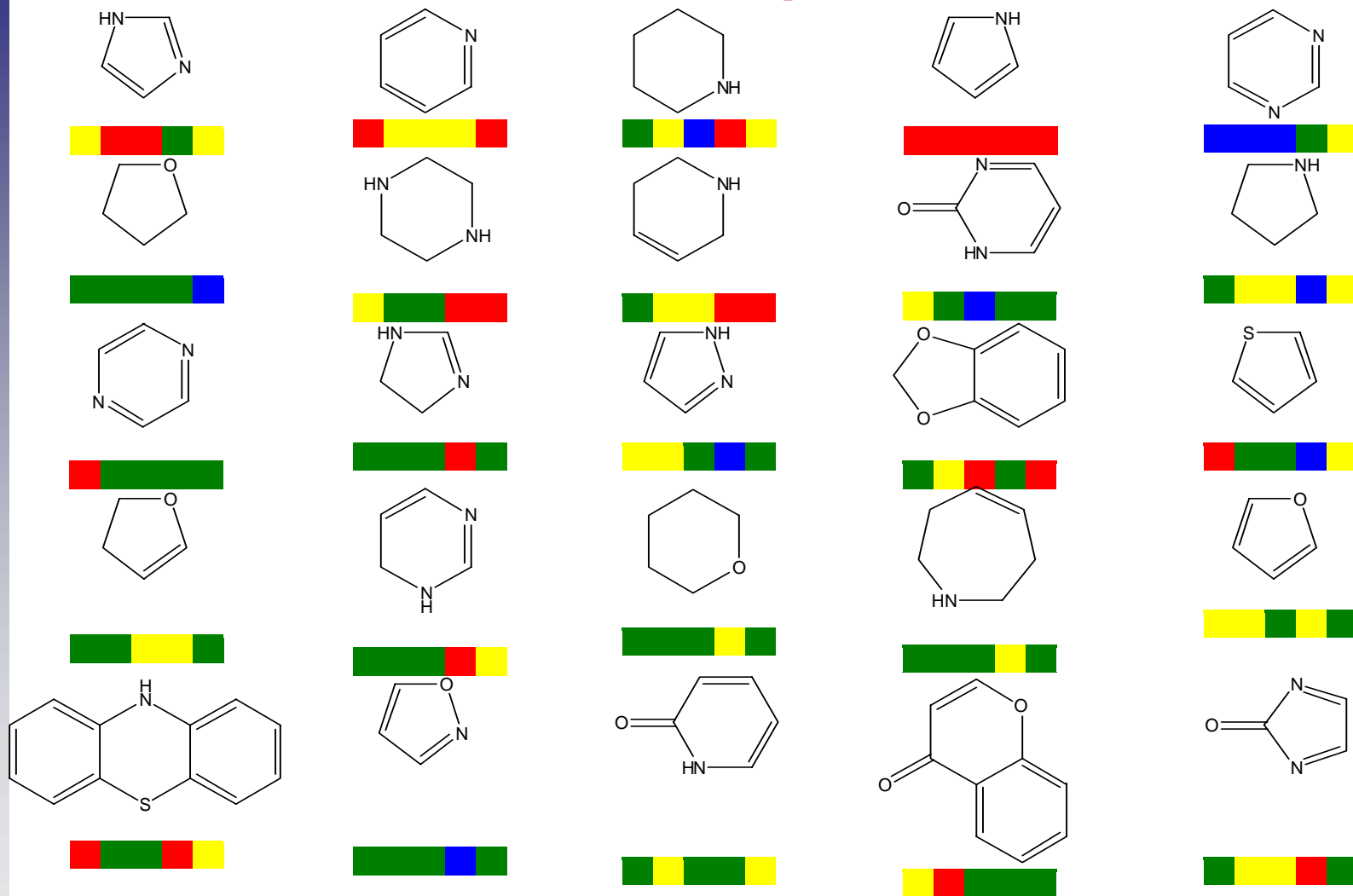


not in drugs

CYP Inhibiton

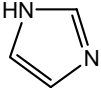
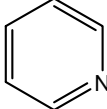
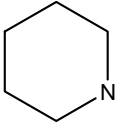
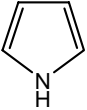
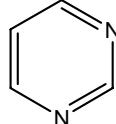

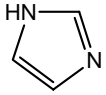
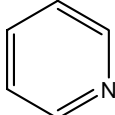
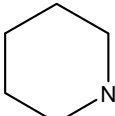
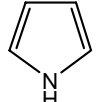
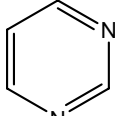

- CYP 1A2, 2C9, 2C19, 2D6 and 3A4
- LOPAC library
 - 1260 compounds, pharmacological probes
 - Scaffold distribution similar to drug sets
 - Diverse
- OR determined for each scaffold
 - Against entire dataset
 - Against other scaffolds

Results for Simple Scaffolds



1A2, 2C9, 2C19, 2D6, 3A4;  OR < 1;  OR ~ 1;  OR > 1

Pairwise Comparisons

1A2						
	—	~	↑	~	~	↑
	~	—	↑	~	~	↑
	←	←	—	←	←	~
	~	~	↑	—	↑	↑
	~	~	↑	←	—	↑
	←	←	~	←	←	—

Odds ratios for each pair of scaffolds
 arrows points to scaffolds with lower odds

Alternative Data Mining Techniques

- Substructural Analysis
 - Attuned to non-experts
 - Supplement current methods
- Odds ratio can serve as rule support
 - Not so common in chemoinformatics
 - Confidence intervals for proportions
- Public databases
 - May be limited
 - Do not necessarily overlap with drugs

Altoris, Inc.

www.altoris.com

www.chemapps.com