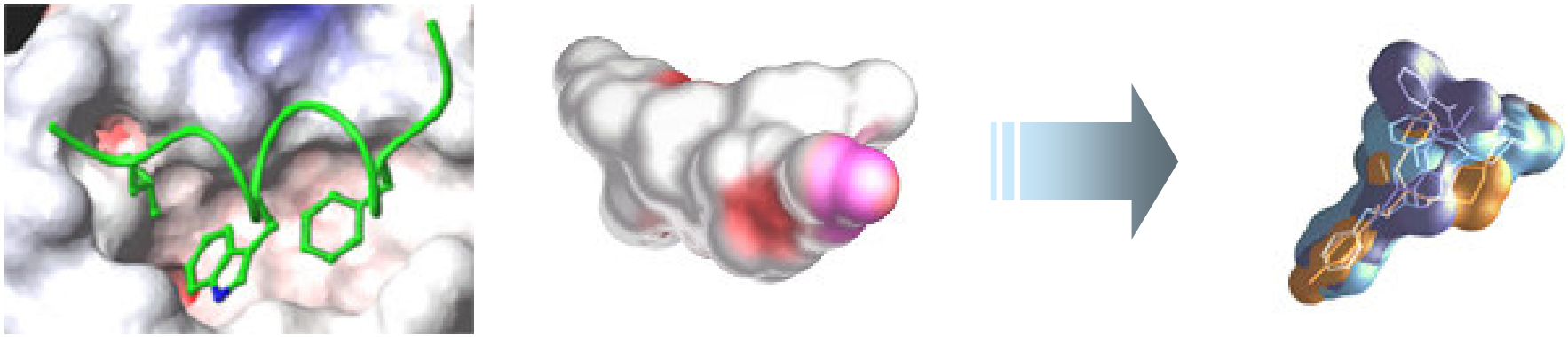
# Introduction

ontochem

**OntoChem®**

- **Providing drug discovery knowledge & small molecules...**
- **Supporting the task of medicinal chemistry**
- **Allows selecting best possible small molecule starting point**
- **From target to leads candidates within a few months**
- **Generating new intellectual property**



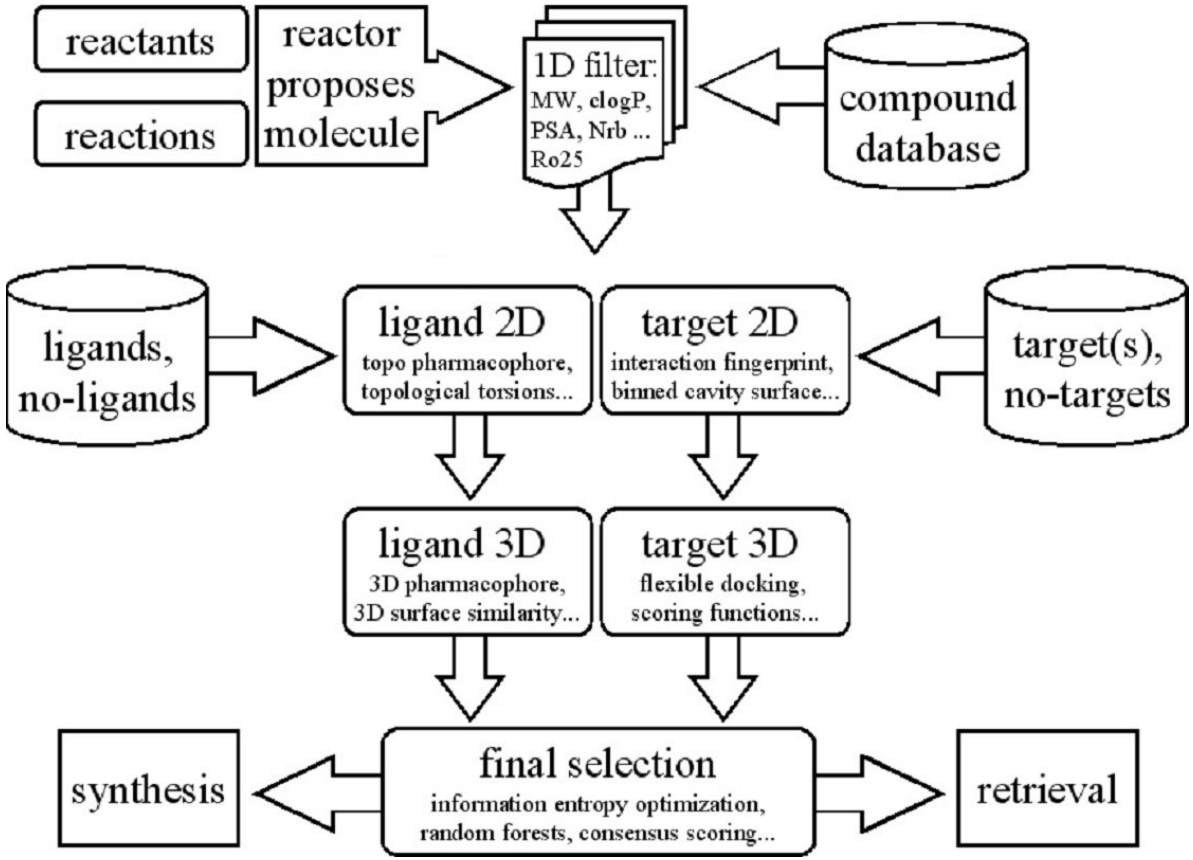
# OntoChem® *drug design process*

ontochem

## Process flow chart

*Design of protein interaction inhibitors using a novel chemo- and bioinformatic platform based on information theory and rational design*

*see*
L Weber, *QSAR & Combinatorial Science,* **2005**, 809-823.

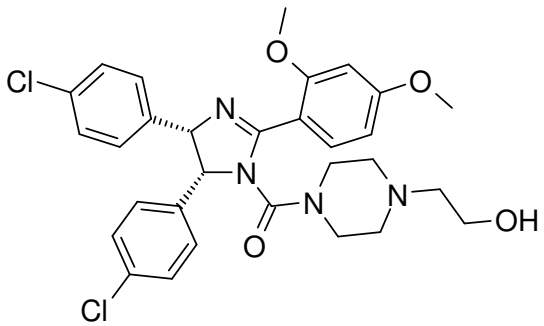


# OntoChem® *example 1*

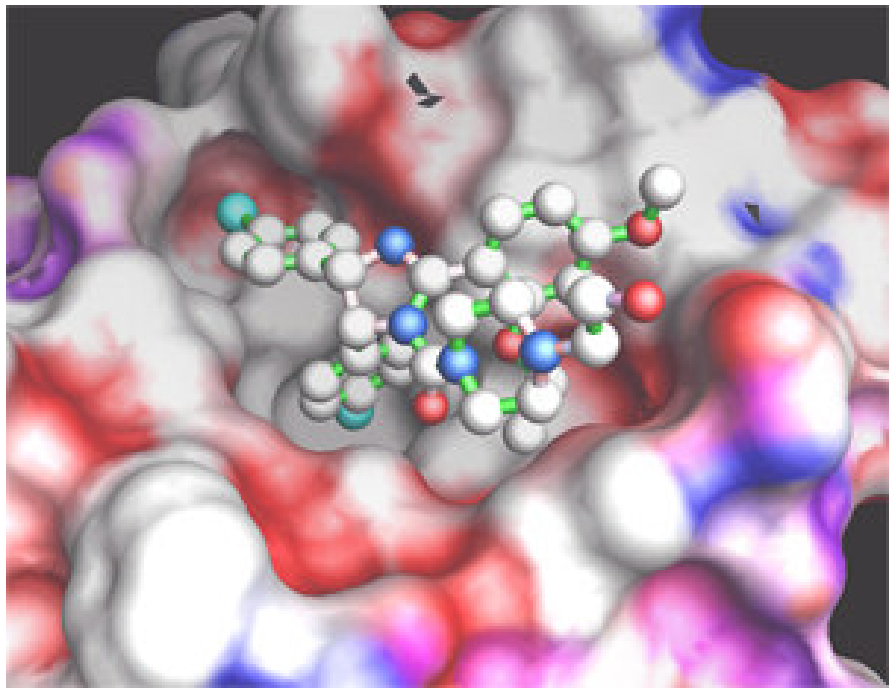
ontochem

## MDM2 - p53 inhibitors discovery project

- A protein-protein interaction with promises in oncology
- Has been a challenge in drug discovery




**Roche “Nutlins”**

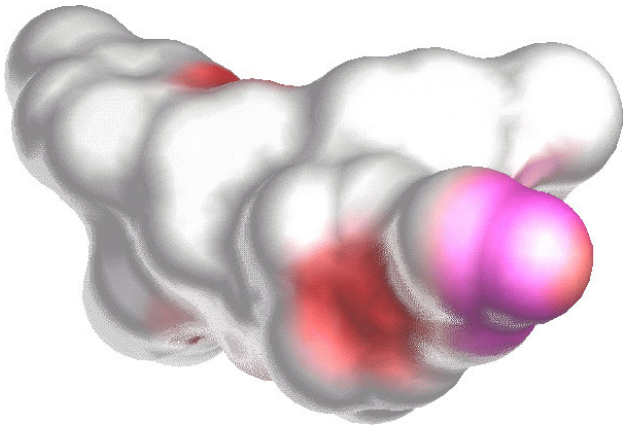


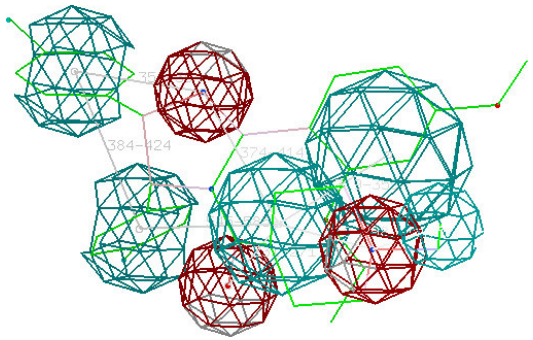
# OntoChem® *example 1*

ontochem

## MDM2 - p53 inhibitors, step 1 - ligand based design

- Generating surface and pharmacophore properties of ligand
- Other criteria: MW<500, Lipinski-rules
- Search terabyte server for sub-set that satisfy criteria





- 2D surface property scoring of selected molecules with input ligand

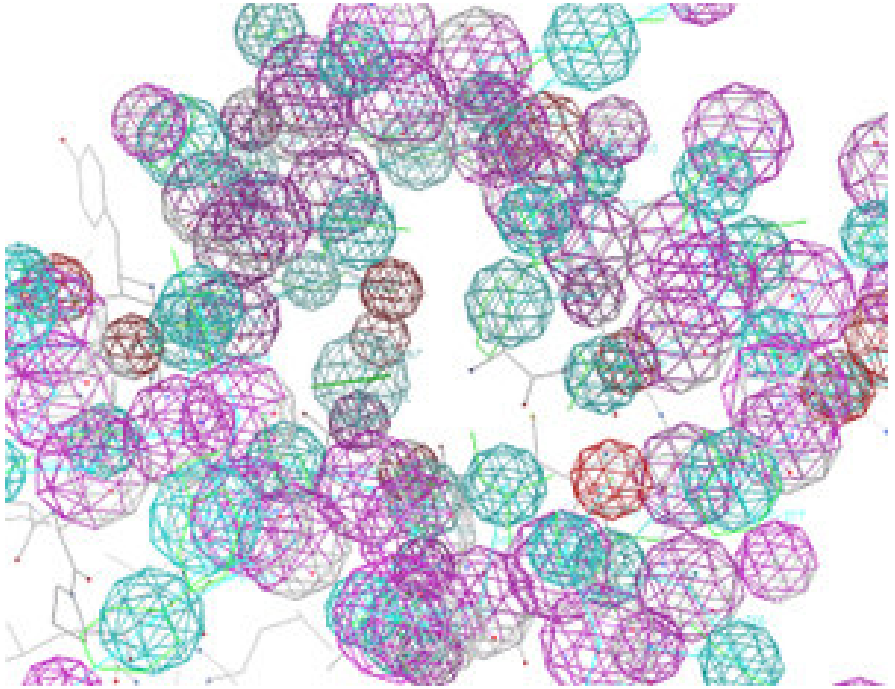
# OntoChem® *example 1*

ontochem

## MDM2 - p53 inhibitors, step 2

- Docking of selected molecules into protein pharmacophore
- Docking best molecules into protein
- Scoring (orthogonal fusion)

  ***Success rate: 0.000009%***
  ***(9 out of 100 million)***
- Resulting into ranked hit list of molecule proposals to be synthesized by wet-lab chemistry



**Found molecules (2 scaffold series) that are active and selective**

# OntoChem® *example 1*

ontochem

## Inhibitor series properties

- MW < 400
- Polar surface area PSA= 41
- H-bond donor 0, acceptor 5
- cLogP 4
- Rotatable bonds 3
- Synthesis via a two step procedure, one step is a MCR
- Straightforward MedChem optimization possible

- Filed two patents

## Roche Nutlin series properties

- MW 583.5
- Polar surface area PSA= 78
- H-bond donor 2, acceptor 8
- cLogP 5
- Rotatable bonds 7
- Synthesis via 12 sequential steps

# OntoChem® *terabyte server*

ontochem

## **Numbers** (…are not everything but may help):

- **New search technology**

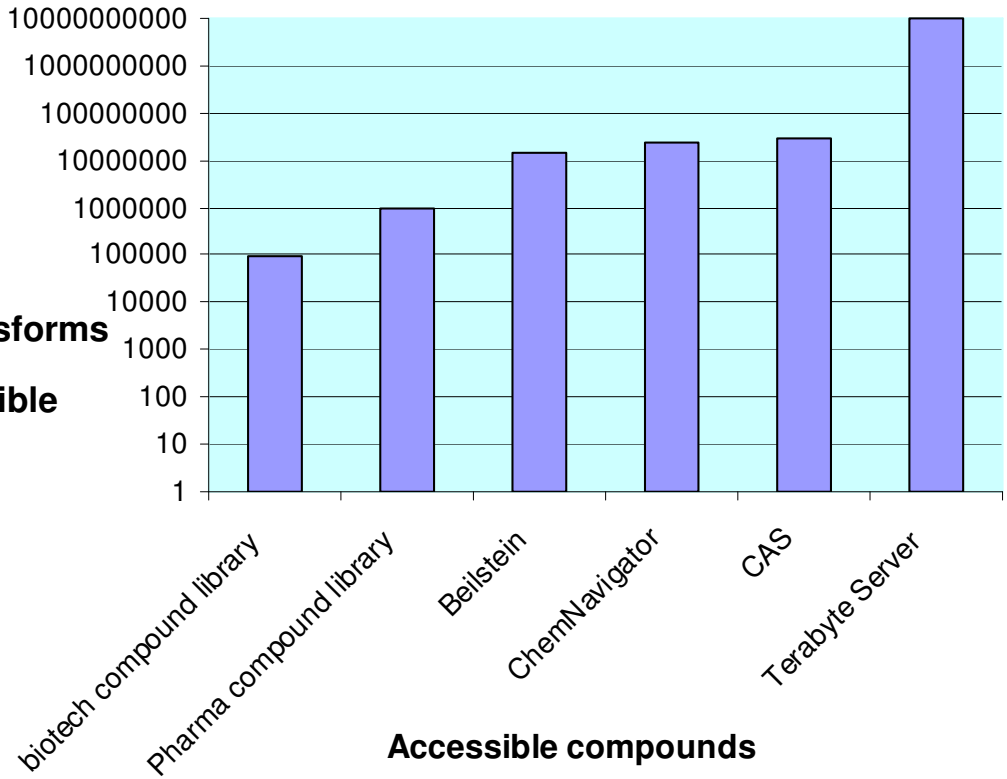
## Innovative Content:

- **MCRs are the most efficient way to construct molecules**

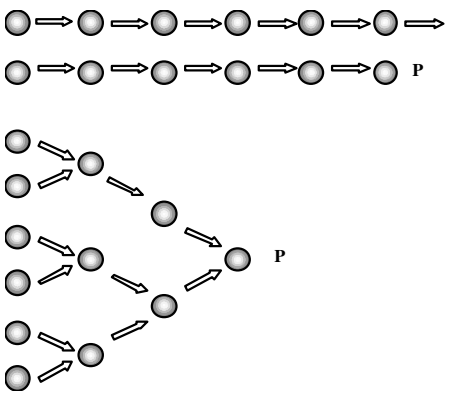
**Products by worked >4000 MCRs and 100 classical transforms**

**From proprietary starting materials or otherwise accessible**

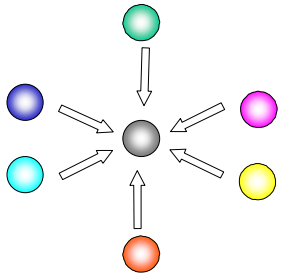
**Preselected by MW < 500 and fuzzy Lipinski filter**




**Classical chemistry**




**MCR chemistry**



6-CR

# OntoChem® *CARD + DayCart access*

ontochem

**CARD + DayCart provides straightforward "everywhere" access to**

- Reaction + transform database (currently >4000 on file)
  - Basis for generating synthetically accessible compounds
  - Updated with worked reactions - scope and limitations
- Project databases for collaborative projects
  - Integrating biological data
  - Calculated physico-chemical data
- Terabyte server compound database ($10^{10}$ compounds on file)
  - Selected for Lipinski rules
  - Accessible by 1-3 steps of straightforward chemistry
  - Basis for HT *in silico* screening

# OntoChem® *CARD + DayCart access*

ontochem

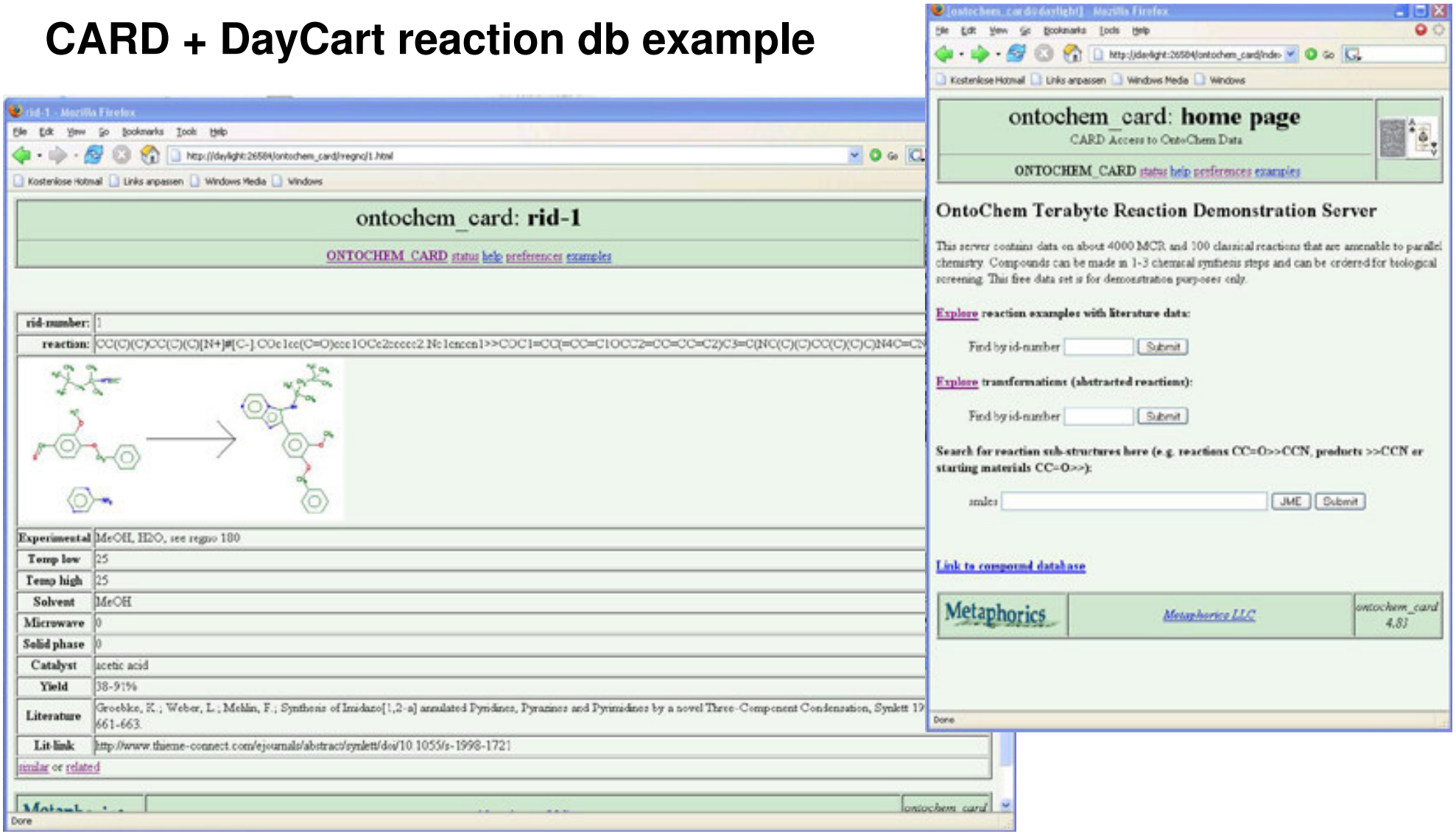
## CARD + DayCart advantages

- No need for large centralized or distributed infrastructure
- Integration with the CABINET federation of other databases
- Highest flexibility
- Easy to setup and run

# OntoChem® *CARD + DayCart access*

ontochem

## CARD + DayCart reaction db example



# Terabyte Server Search Methods: Chemical Similarity *today's problem*

ontochem

E.g., ***1a*** is calculated most similar to ***2b,*** using Tanimoto similarity of bitstrings = ***wrong!***

(M. Stahl et al., *J. Med. Chem.* **2005**, *48*, 4358)

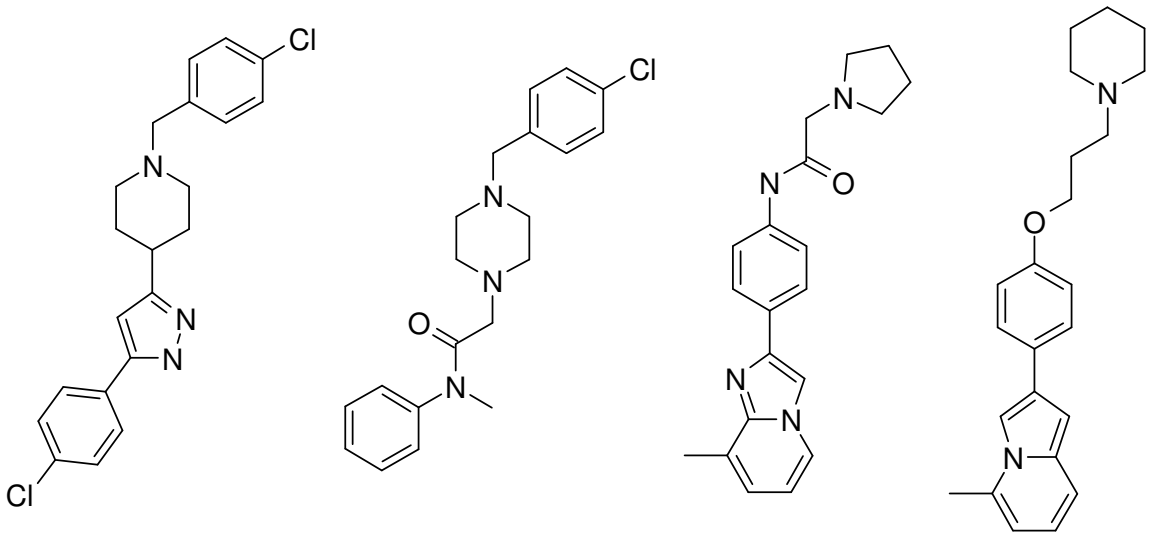
**Daylight fingerprints**

| | 1a | 1b | 2a | 2b |
|---|---|---|---|---|
| 1a: | 1.00 | 0.27 | 0.26 | 0.32 |
| 1b: | 0.27 | 1.00 | 0.31 | 0.19 |
| 2a: | 0.26 | 0.31 | 1.00 | 0.37 |
| 2b: | 0.32 | 0.19 | 0.37 | 1.00 |

**MACCS keys**

| | 1a | 1b | 2a | 2b |
|---|---|---|---|---|
| 1a: | 1.00 | 0.56 | 0.53 | 0.54 |
| 1b: | 0.56 | 1.00 | 0.64 | 0.50 |
| 2a: | 0.53 | 0.64 | 1.00 | 0.62 |
| 2b: | 0.54 | 0.50 | 0.62 | 1.00 |

***1a*** ***1b*** ***2a*** ***2b***




**Dopamine D4 antagonists**

**Histamine H3 receptor ligands**

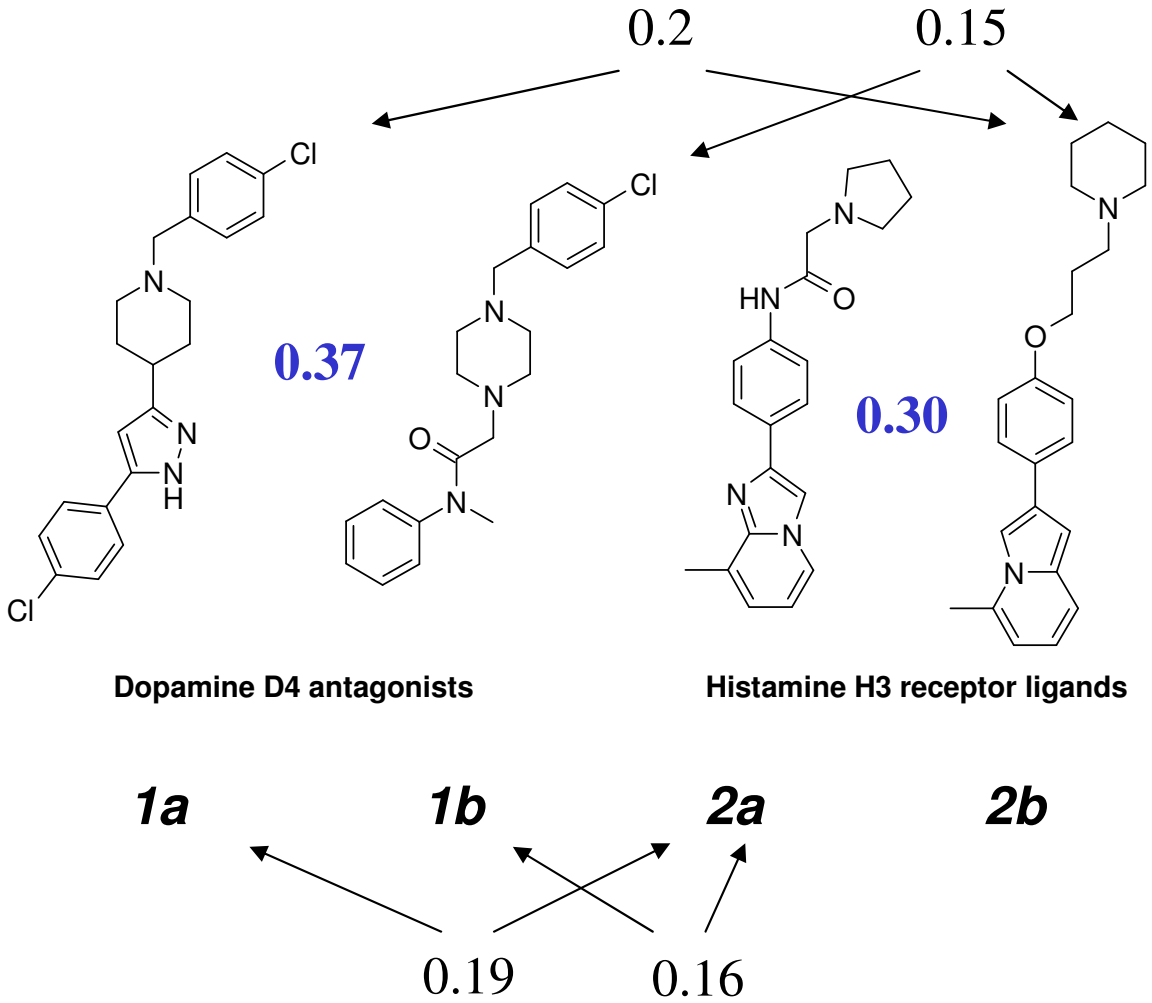
# Topological Torsion *validation*

ontochem

**TT similarity allows better classification of compounds than by other known methods** *(from Nilakatan 1987 to Sheridan 2004)*

```
TT - Tanimoto

     1a    1b    2a    2b
1a: 1.00 0.37 0.19 0.20
1b: 0.37 1.00 0.16 0.15
2a: 0.19 0.16 1.00 0.30
2b: 0.20 0.15 0.30 1.00
TT - Dice
1a: 1.00 0.54 0.31 0.33
1b: 0.54 1.00 0.27 0.27
2a: 0.31 0.27 1.00 0.47
2b: 0.33 0.27 0.47 1.00
```

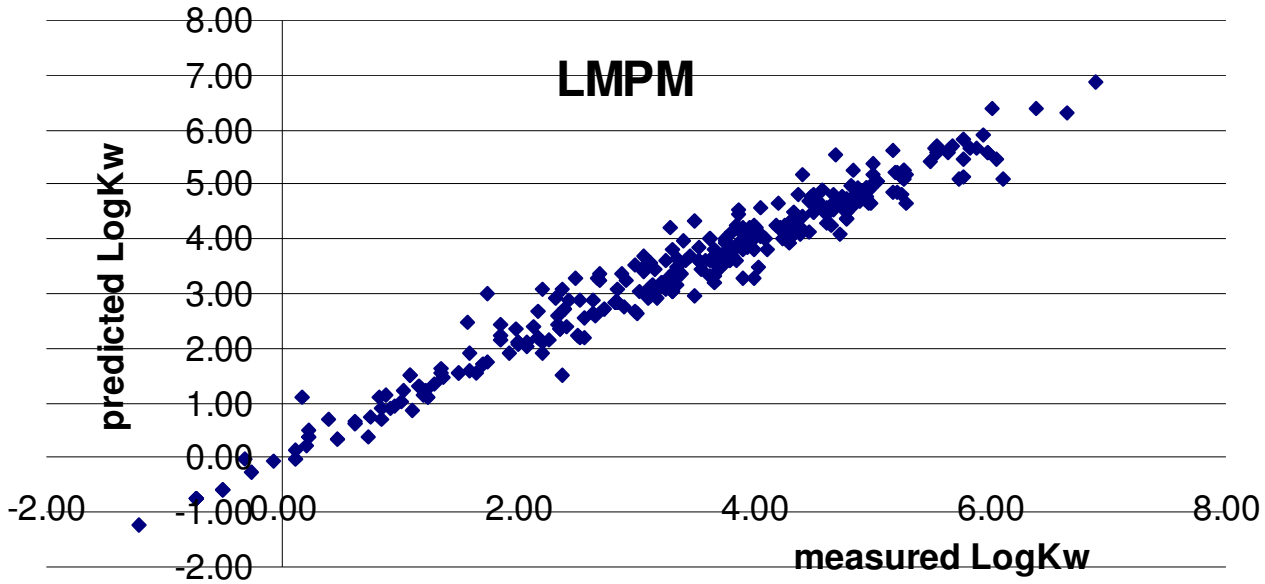


# TT *validation*

ontochem

**Example for the linear multi-pharmacophore model (LMPM) algorithm, predicting the $LogK_w$ for 336 GPCR compounds (T. Oprea) with different scaffolds and active for different GPCRs.**

**Note that clogP's are typically calculated by commercially available, dedicated algorithms, however, the good fit achieved with LMPM demonstrates the universality of the method.**

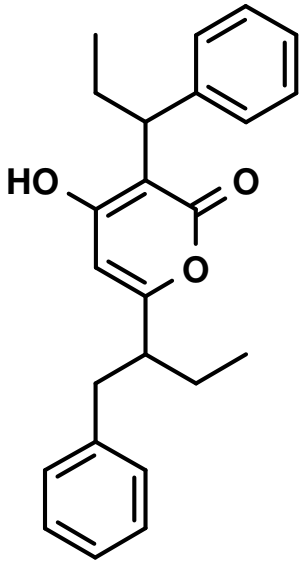


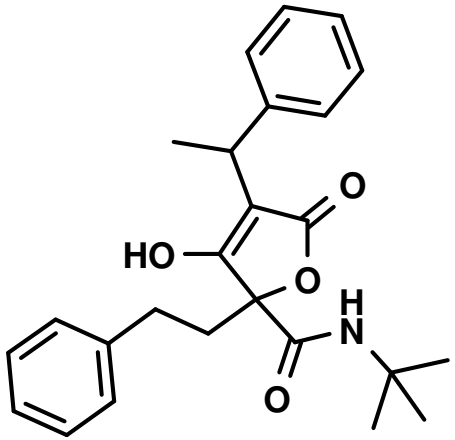
# TT *validation*

ontochem

**HIV protease inhibitor PNU-96988 and a new inhibitor**

- **Daylight Fingerprint similarity 0.16**
- **TT Tanimoto similarity 0.302**



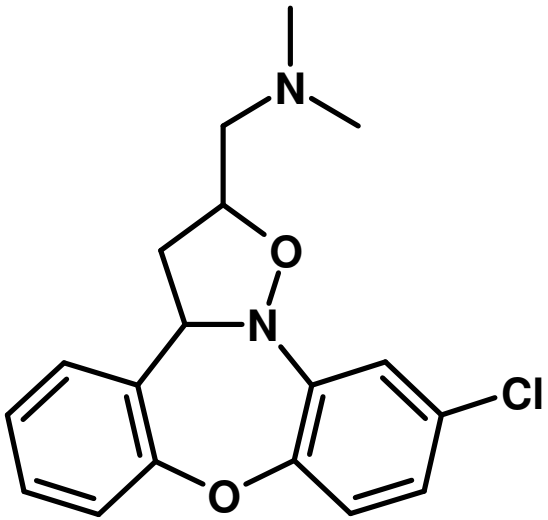
# TT *validation*

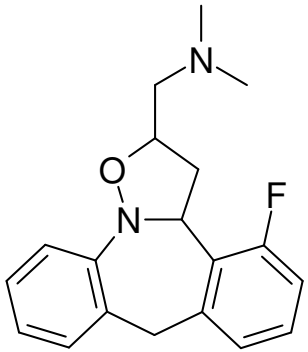
ontochem

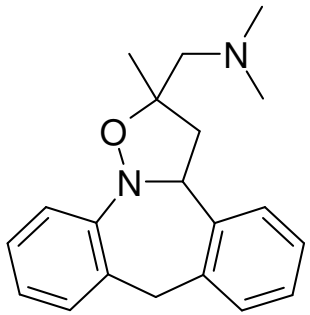
## Tudor Oprea dataset 341 GPCR ligands

- average TT  Tanimoto similarity 0.131
- 49 have TT similarity > 0.2, 34 *(!)* are 5HT2A ligands…from 50 total

Chiral



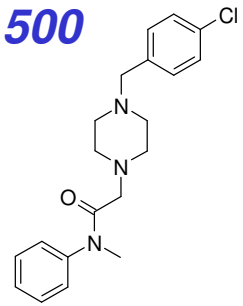

**Not ligands but very similar**

# TT *diversity design*

ontochem

**Shannon information entropy:**

- **Introduced by J Bajorath to drug design**
- **Information is always a measure of the decrease of uncertainty**
- **Medicinal chemistry can be regarded as a process to obtain *maximal information* from an input (compounds) and an output (test results)**
- **Maximal information can be obtained if number of information channels is maximized: diverse chemical structures = different TT's**
- **Defined here as** $H = -\Sigma(p_i/n)*log_2(p_i/n)$
  - *i-state of n total states, $p_i$-occupation of i-state*
  - *TT's may be used as state descriptors for small molecules*
  - *approx* $n = 100$ *for small molecule MW < 500*
  - *approx* $H_{max} = 6.64$ *and e.g.* $H_{benzene} = 0$
  - $H = 4.57$ *(32 different TT's,* $n = 92$*) for*



# TT *diversity design*

ontochem

**Information entropy diversity design method:**

- ***Maximize number of different TT's in a molecule***
- ***Minimize number of equal TT's in different molecules***

**What is the gain in $H$ by adding a new molecule to the library?**

- **In regard to new and different TT's**

**Max-min diversity design:**

- **Optimally diverse library - minimize TT similarity matrix:**

```
1.00 0.00 0.00 0.00
0.00 1.00 0.00 0.00   4 molecules do not share common ToTo's
0.00 0.00 1.00 0.00
0.00 0.00 0.00 1.00
```

- **Optimally diverse library - maximize $H$ entropy of each molecule:**

```
4.57 4.57 4.57 4.57
```

# Library Information Entropy

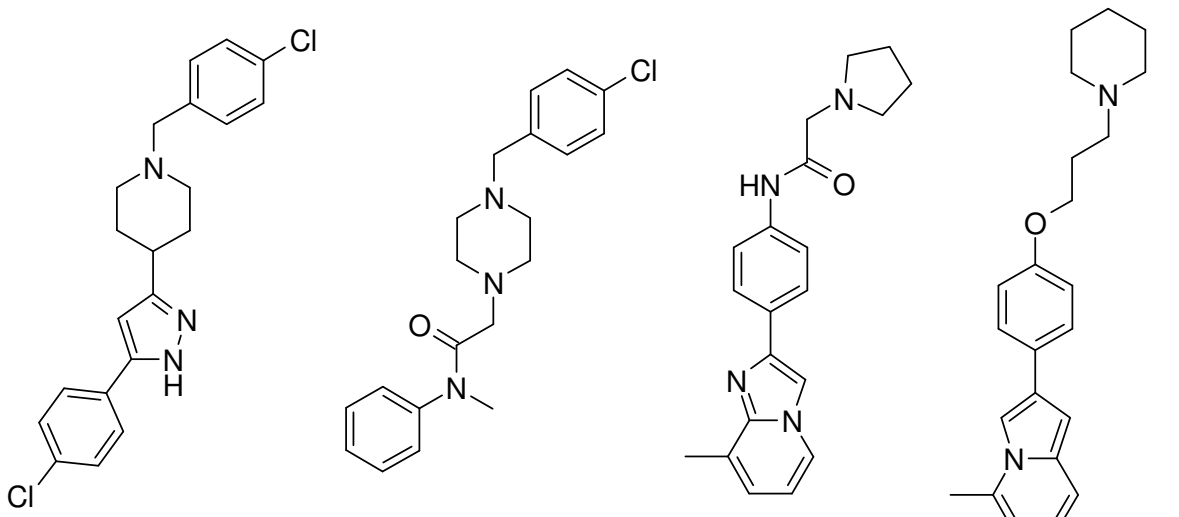
ontochem

## TT allows straightforward information entropy calculation

TT

| | 1a | 1b | 2a | 2b |
|---|---|---|---|---|
| 1a: | 1.00 | 0.37 | 0.19 | 0.20 |
| 1b: | 0.37 | 1.00 | 0.16 | 0.15 |
| 2a: | 0.19 | 0.16 | 1.00 | 0.30 |
| 2b: | 0.20 | 0.15 | 0.30 | 1.00 |

$H_{molecule}$ = 5.18 4.57 5.62 5.53




Dopamine D4 antagonists

Histamine H3 receptor ligands

$H_{library}$ = 6.30 (n = 400, only 142 different), maximum = 8.64

# Reaction Classification

ontochem

**Using TT's - similar to the InfoChem's CLASSIFY**

**Transform TT into SMIRKS...**

# TT Utility *medchem selector*

ontochem

## Medicinal chemistry applications of TT similarity analysis

- Diversity Analysis of corporate/vendor chemical library
  - Analyze information content of a library
  - Gap analysis - synthesize molecules to fill the gap
  - Maximize diversity
- SAR – prediction of small molecules
  - Target class specific fingerprints (e.g. using known kinase inhibitors)
  - Target specific fingerprints (e.g. using known Aurora inhibitors)
- Property Prediction – allows efficient substructure design
  - Exclusion of molecules
  - Activity
  - cLogP
  - ADMET
  - ……
  - Selection of most promising candidates *in-silico*