

# UK QSAR Society

Virtual screening: The king (quite often) has no clothes

Robert Glen  
Andreas Bender



UNIVERSITY OF  
CAMBRIDGE



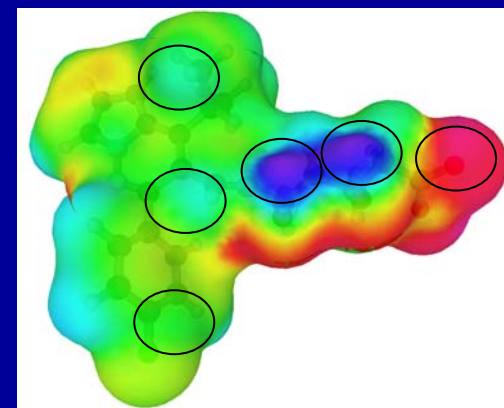
# Virtual screening

- Mostly used in pharmaceutical research to generate starting points for experimental screening
- 'In Silico' approach – in the virtual world of the computer
- Expectation...
  - Save time and effort, increase efficiency, lower cost, be comprehensive, discover novelty...
- But, we live in the real world so, we need to be circumspect and be careful about how useful or applicable these methods really are
- We will describe some of our recent approaches, based on fingerprints in both 2D and 3D, what they find and how reliable (or otherwise) they seem to be.

# Applications...

Tree structured (Circular) fingerprints (2D and 3D) - describe 'patches' on (in) molecules

These local regions can be used to describe different molecular properties. Therefore, properties that depend on a collection of 'environments' e.g. ligand/protein binding can reveal which environments appear to be related to the property.



pKa prediction

Metabolism prediction

Toxicity prediction

Similarity

Virtual screening

We have tried out these fingerprints in a variety of applications

Move to 3D fingerprints

pharmacophore perception

protein binding features

Our first interest was to use tree structured fingerprints to describe the environment around an ionizable center (an atom environment) – predict a pKa

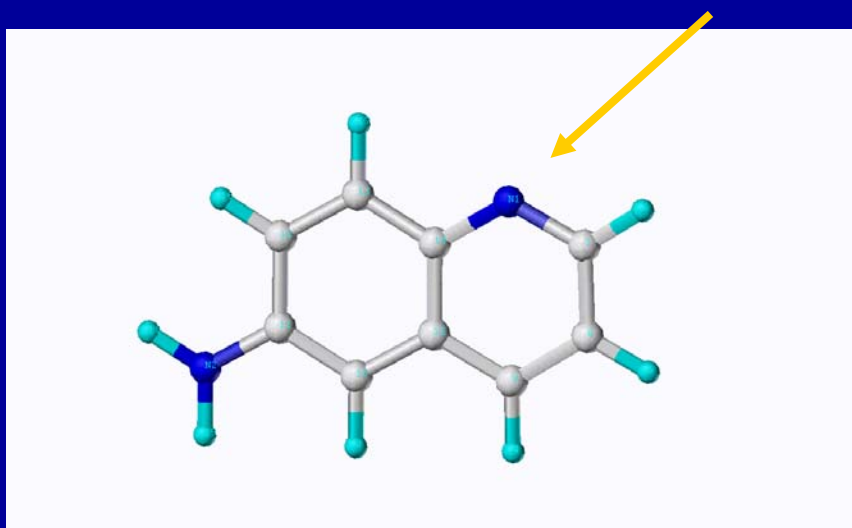
## Method

- Tabulate many reliable pKa measurements
- Describe the environment around ionizable centers
- Use partial least squares to create a predictive model
- Test model with cross validation

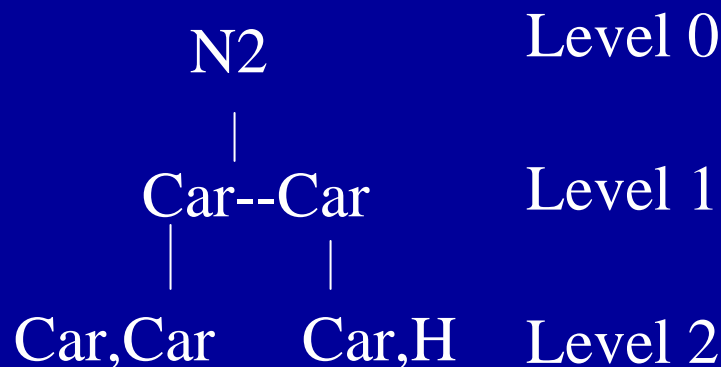
# A tree structured (Circular) fingerprint

- E.g. 6-aminoquinoline

Measured 5.7  
predicted 5.4



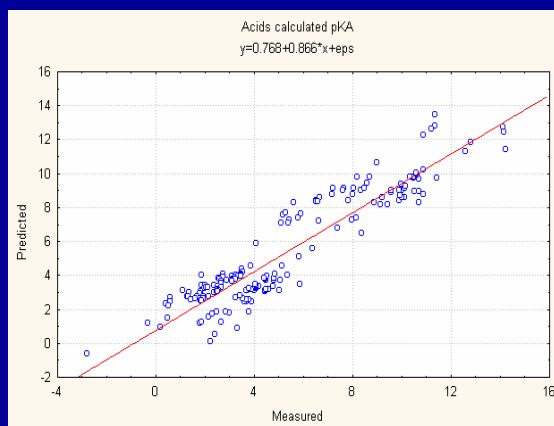
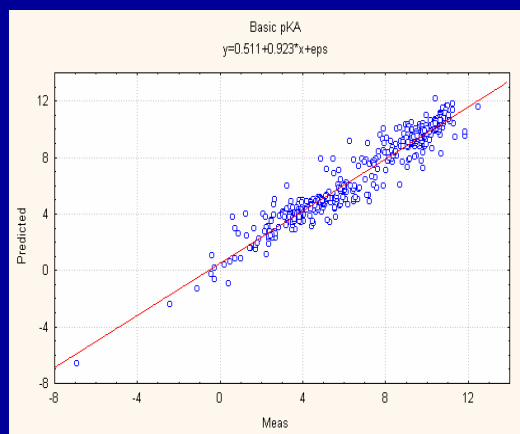
Start with interesting atom  
find connections  
find connections to connections  
create a tree down to 5 levels  
'bin' the atom types for each level  
create a 'fingerprint' for this atom



String contains a bin for each required atom type at each level,  
the number of atom types is accumulated to form the string - 56 bins

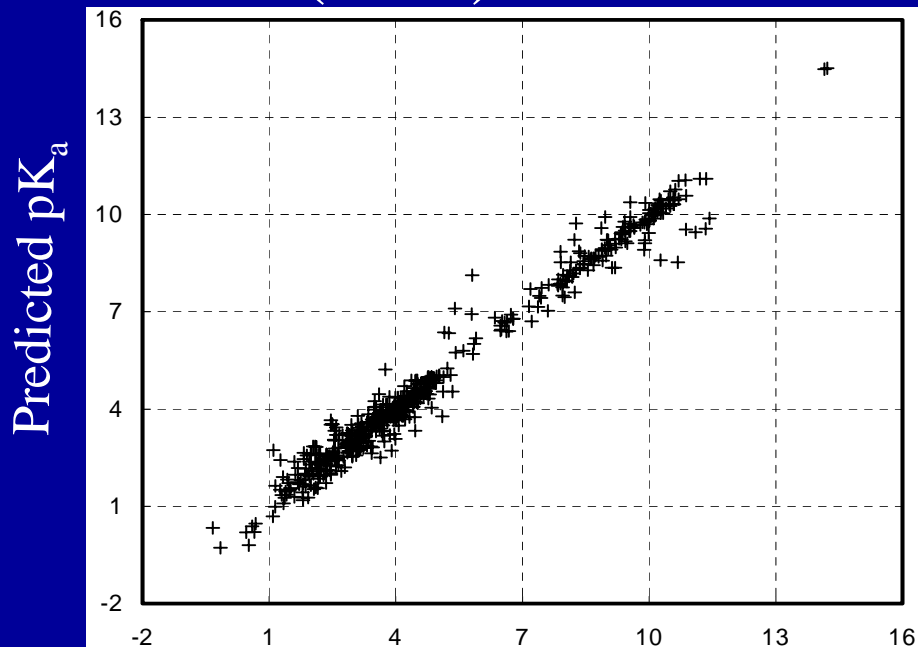
# Using the data

- 56 bins used to cover all the possibilities
- Used pls (partial least squares) to create a model
- $pK_a = pK_c^0 + \sum a_i x_i + \sum g_j y_j + \sum q_k z_k \dots$
- Used cross validation to validate the model
- *Novel methods for the prediction of pKa, logP and logD*, Xing L. and Glen R.C.. J. Chem. Inf. Comput. Sci.; **2002**; 42(4); 796-805



- Next, refined the model to improve accuracy

# pKa of acids (625)



$R^2=0.98$

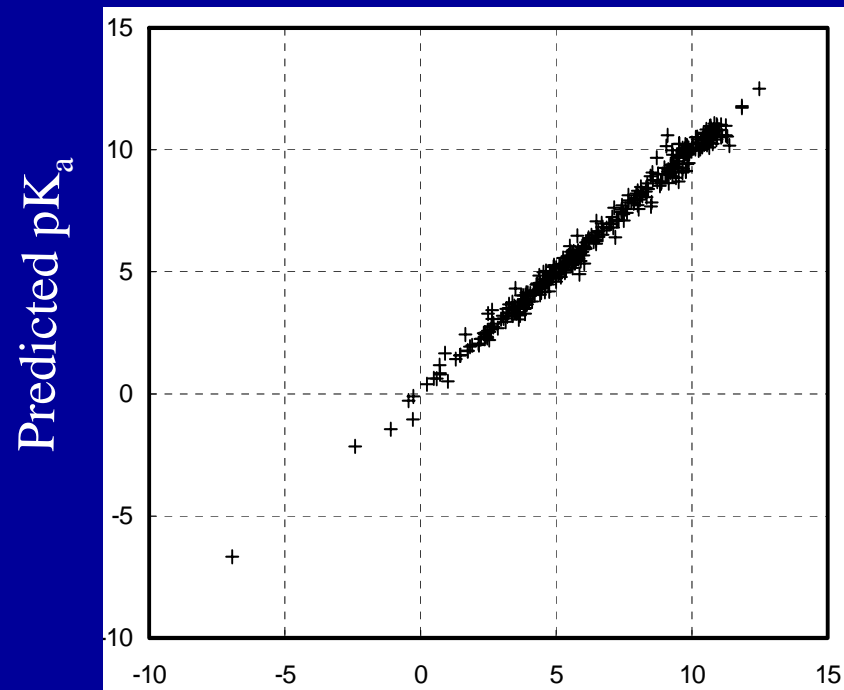
Measured pK<sub>a</sub>

Std.Err.=0.405

N=625

$Q^2=0.92$

# pKa of bases (412)



Measured pK<sub>a</sub>

$R^2=0.99$

Std.Err.=0.302

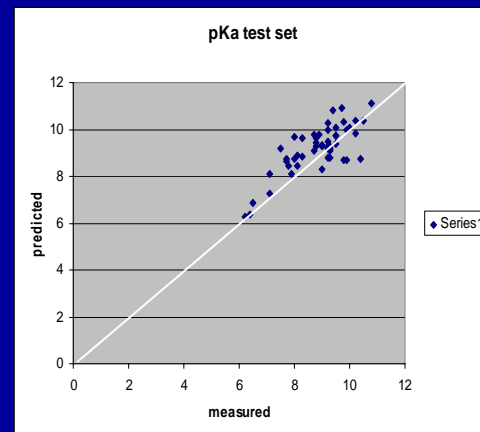
N=412

$Q^2=0.95$

Looks over-trained, but the results are promising

# Conclusions

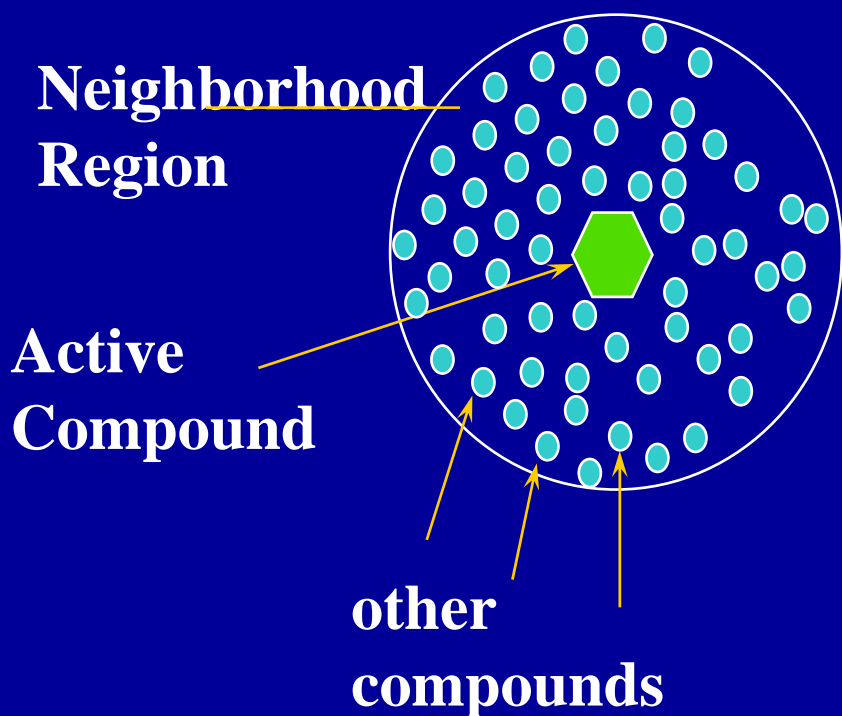
- Surprisingly good results - fast
- Predictive for most pK's
- Useful in biological setting in estimating Pharmacokinetics, active species, metabolism etc.
- Predicts for all types - sometimes get odd results though, if outside parameter set or the 'atom types' are miss-set
- Can apply these fingerprints to other problems e.g. molecular similarity



Novel methods for the prediction of pKa, logP and logD, Xing L. and Glen R.C.. J. Chem. Inf. Comput. Sci.; 2002; 42(4); 796-805

Predicting pKa by Molecular Tree Structured Fingerprints and PLS. Xing L.,Glen R. C. and Clark, R. D. J. Chem. Inf. Comput. Sci. 2003, 43(3), 870

# Database searching using a similarity approach with circular fingerprints— how good can it be and how far can we trust the results ?



If the molecular descriptors are valid ...

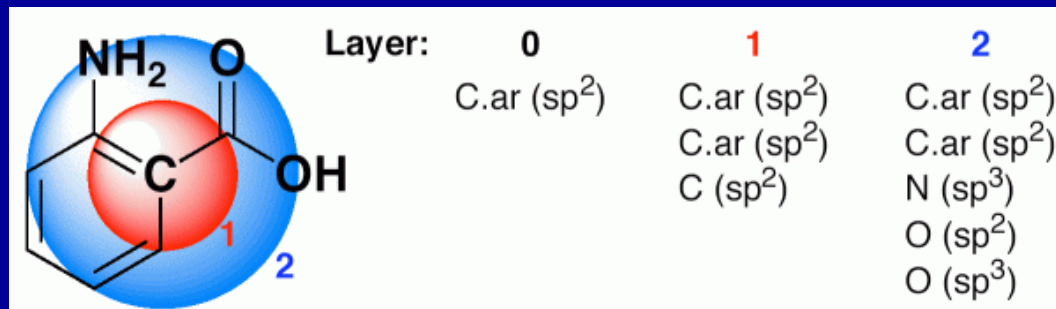
the activity of a Compound is shared by *most* other compounds within its Neighborhood Region

i.e. neighbors of a bioactive compound have a higher probability of behaving in a 'similar' bioactive way

Molecular similarity: a key technique in molecular informatics. Organic and Biomolecular Chemistry perspective article. R. C. Glen and A. Bender, *Org. Biomol. Chem.* 2004, 2, 3204 - 3218.

# MOLPRINT 2D

- 1. Fingerprint features



- 2. Information-Gain Feature Selection

$$I = S - \sum_v \frac{|S_v|}{|S|} S_v \quad S = -\sum p \log_2 p$$

- 3. Naïve Bayesian Classifier

$$\frac{P(CL_1 | F)}{P(CL_2 | F)} = \frac{P(CL_1)}{P(CL_2)} \prod_i \frac{P(f_i | CL_1)}{P(f_i | CL_2)}$$

# 2. Information-Gain Based Feature Selection

- There may be many features – only some may be 'important' in identifying the active compounds
- We wish to select the important features
- To do this we calculate the entropy of the data as a whole and for each class.
- This is used to select those features with the highest discrimination, e.g. active or inactive or toxic and non-toxic molecules

$$S = - \sum p \log_2 p$$

$$I = S - \sum_v \frac{|S_v|}{|S|} S_v$$

# 3. Classification

- The next step is to identify which molecules belong to which class.
- To do this we use a Naïve Bayesian Classifier using the features (atom environments) we have identified as being important.

# 3. Naive Bayesian Classifier

- Include all selected features  $f_i$  in calculation of

$$\frac{P(CL_1 | F)}{P(CL_2 | F)} = \frac{P(CL_1)}{P(CL_2)} \prod_i \frac{P(f_i | CL_1)}{P(f_i | CL_2)}$$

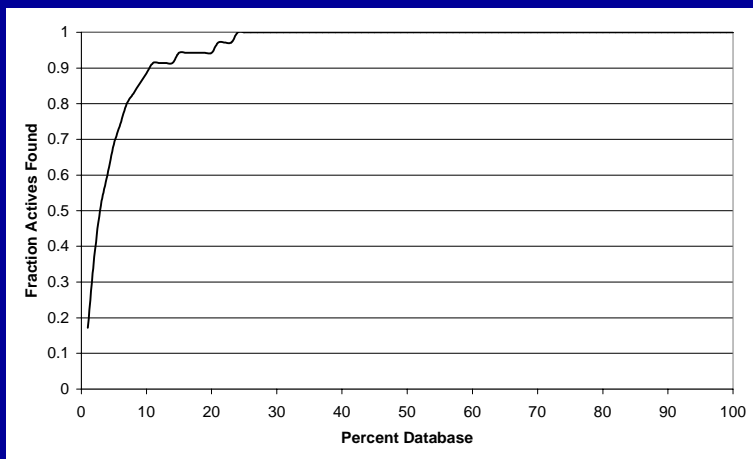
- Ratio  $> 1$ : Class membership 1
- Ratio  $< 1$ : Class membership 2
- $F$ : feature vector
- $f_i$ : feature elements

# Virtual screening examples

- MDDR test run: 957 ligands from MDDR
  - 49 5HT3 Receptor antagonists, 40 Angiotensin Converting Enzyme inhibitors (ACE), 111 HMG-Co-Reductase inhibitors (HMG), 134 PAF antagonists and 49 Thromboxane A2 antagonists (TXA2)
- A) Hit rate among ten nearest neighbours for each molecule
- B) 20-fold Cross Validation, 1-5 Molecules for query generation

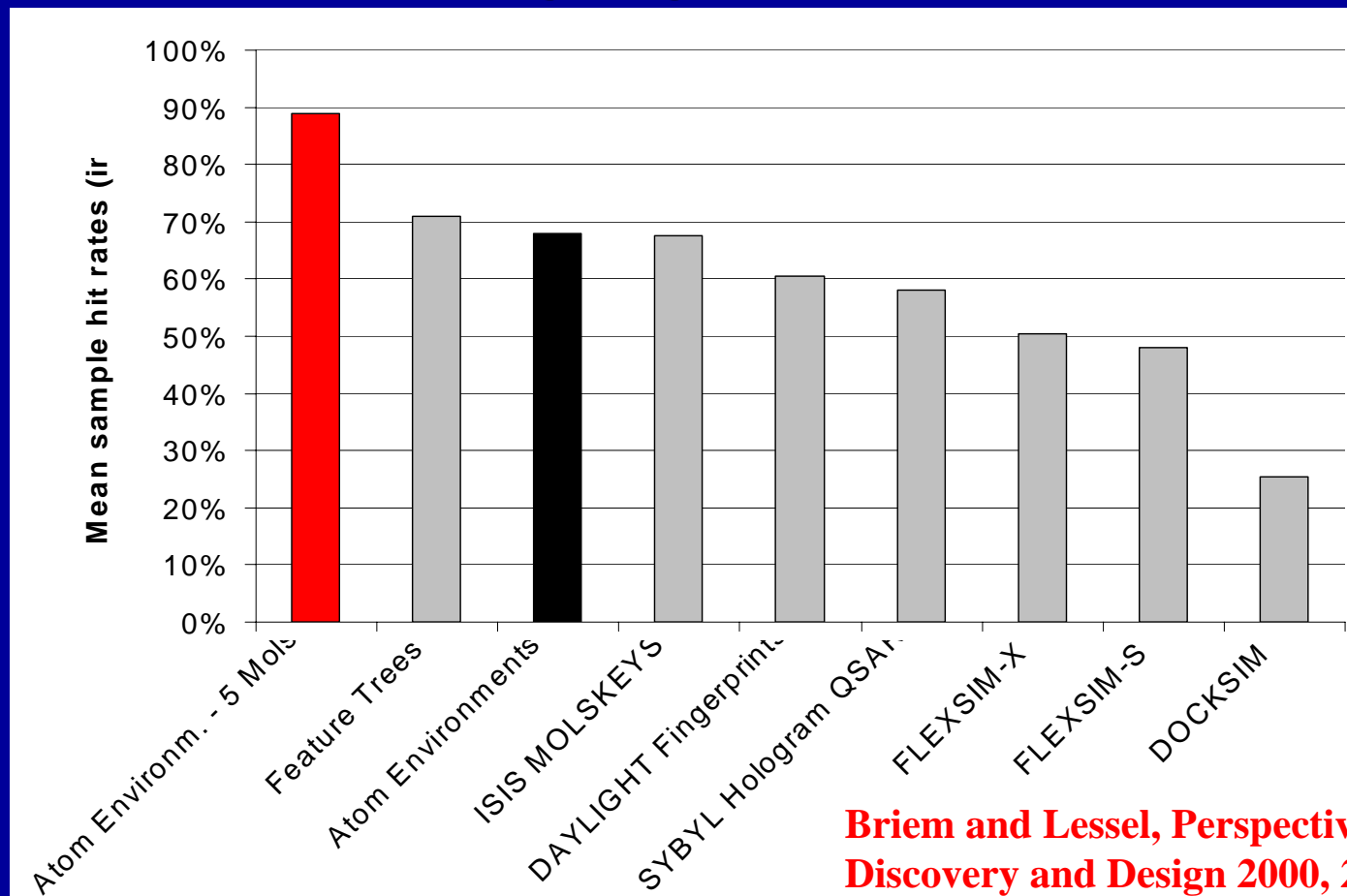
# MDDR database searches

Performance of the Atom Environment Approach, Selecting 20 Features						
Group of Active Compounds	5HT3	ACE	HMG	PAF	TXA2	Overall
Expected Hit Rate for Random Selection	0.50	0.41	1.15	1.39	0.50	0.79
Hit Rate for this Method	5.82	5.85	8.33	7.29	6.47	6.75
Enrichment Factor	11.6	14.3	7.24	5.24	12.9	8.54



e.g. ACE: We found about 80% of the active molecules among the first 10% of the library

# Combining data and search performance - quite encouraging



**Briem and Lessel, Perspectives in Drug  
Discovery and Design 2000, 20, 245-264.**

**Molecular Similarity Searching using Atom Environments, Information-Based Feature  
Selection and a Naïve Bayesian Classifier**

**Andreas Bender, Hamse Y. Mussa and Robert C. Glen, University of Cambridge**

**Stephan Reiling, Aventis Pharmaceuticals**

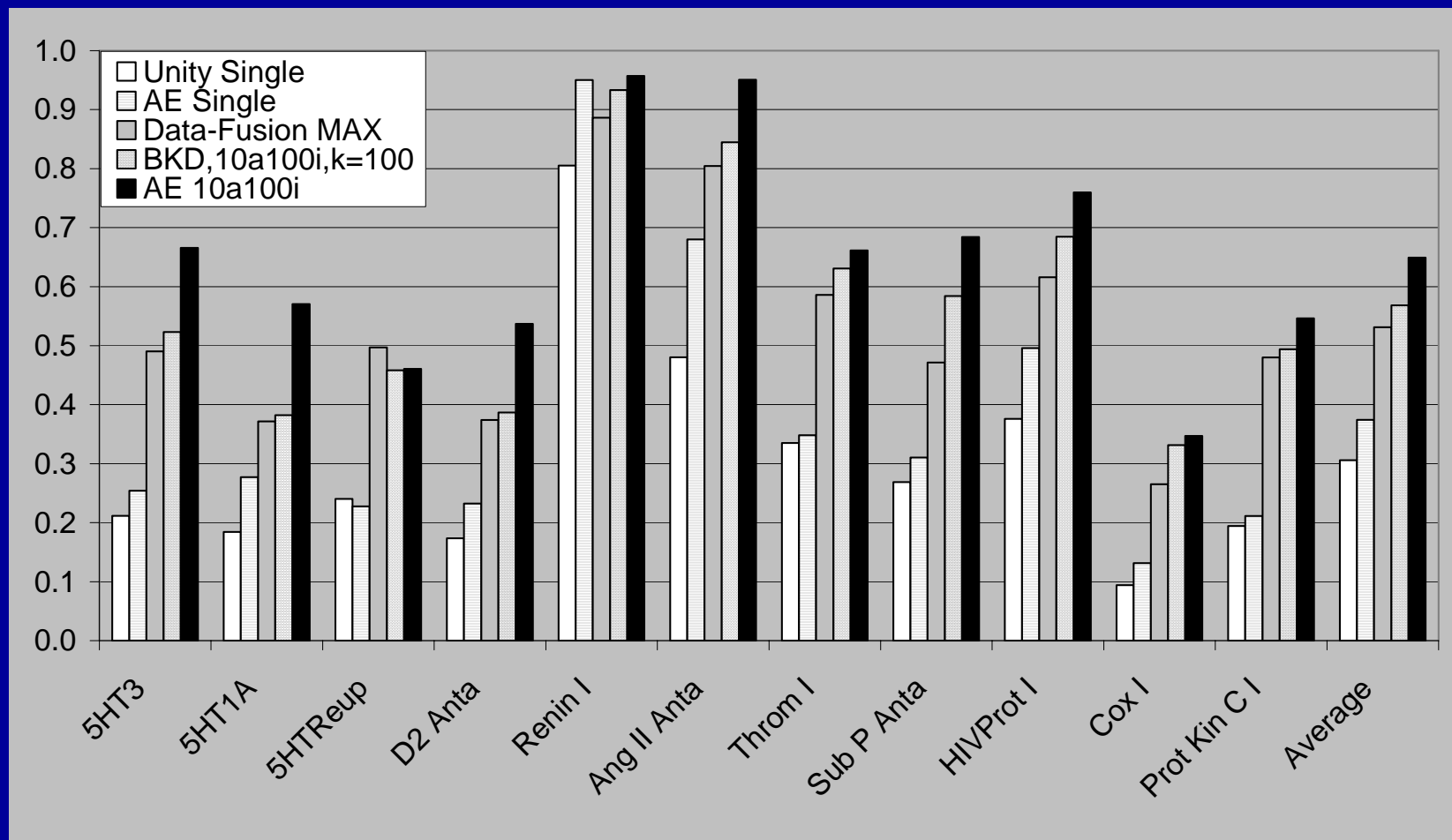
**J. Chem. Inf. Comp. Sci. , 2004; 44(1); 170-178**

# Comparison using Larger Data Set \*

- 102,000 structures from the MDDR
- 11 Sets of Active Compounds, ranging in size from 349 to 1246 entries – large and diverse data set
- Performance Measure: Fraction of Active Structures retrieved in Top 5% of sorted library
- Atom Environments were compared to Unity Fingerprints in Combination with Data Fusion (MAX) and Binary Kernel Discrimination
- In case of Binary Kernel Discrimination and the Bayes Classifier 10 actives and 100 inactives used for training

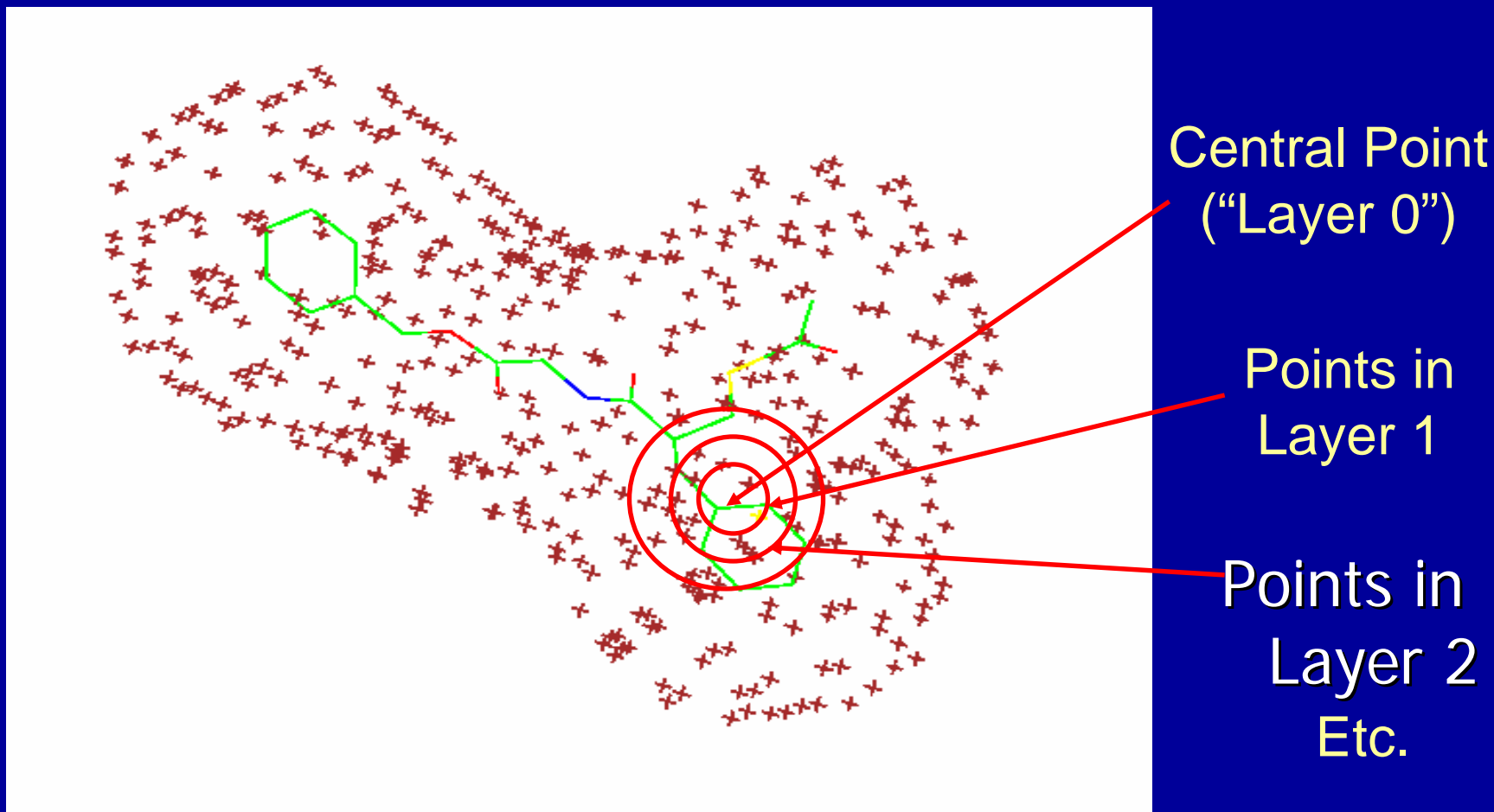
\* Hert J, Willett P, Wilton DJ: Comparison of fingerprint-based methods for virtual screening using multiple bioactive reference structures. J Chem Inf Comput Sci 2004, 44:1177-1185.

# Comparison of Methods – combination of circular fingerprints, feature selection and Bayes Classifier seems to work very well.



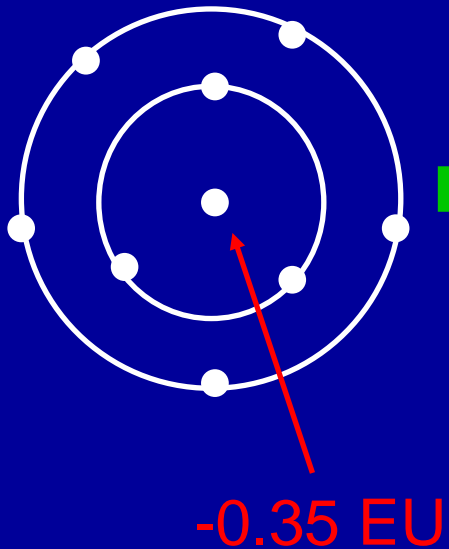
Similarity Searching of Chemical Databases Using Atom Environment Descriptors (MOLPRINT 2D): Evaluation of Performance. Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S.J. *Chem. Inf. Comput. Sci.*, 2004; 44(5); 1708-1718.

# Transformation of similar fingerprints to 3D: Environment around a surface point: solvent accessible surface



# Algorithm

Interaction Energies at Surface Points, one Probe at a time



Binning Scheme



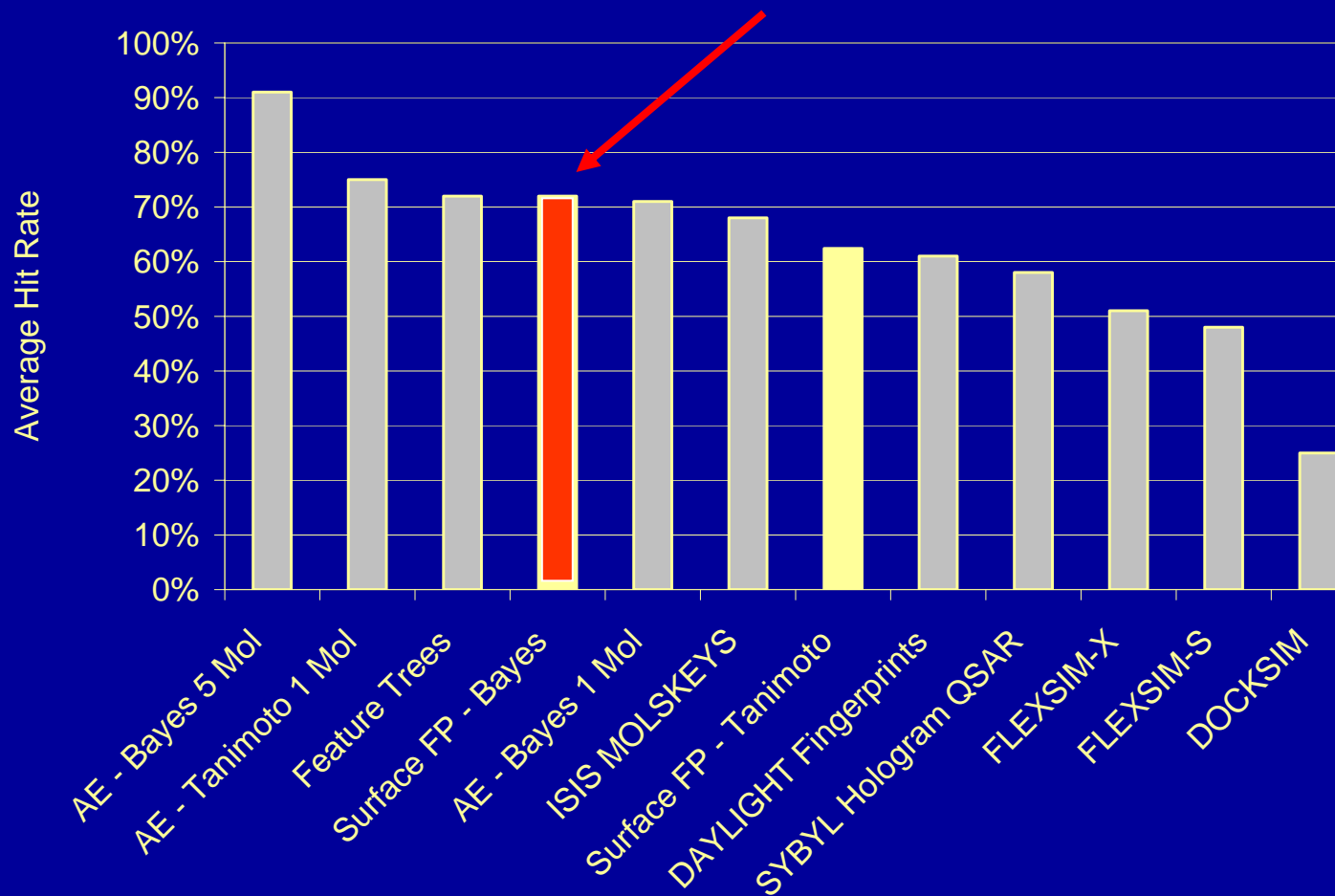
Surface Point Environment

000**1**0000 – 01100010 - 011101100

# Algorithm Flow

Step	Program used	Parameters
Generation of 3D coordinates	Concord	
Calculation of Surface Points	msms	Sphere radius, probe size, triangulation density
Calculation of Interaction Energies	GRID	Probe (and various others)
Transformation of interaction energies into descriptors	Perl script	Binning, number of bins, threshold levels

# Surface Environments – comparison with 2D and other methods – not too bad



## But, are the results sensitive to Conformational Variance ?

- MDDR Dataset (5HT3, ACE, HMG, PAF, TXA2)
- 10 Randomly selected compounds each
- 10 Conformations generated by GA search with large window ( $10^\circ$  for rigid 5HT3,  $100^\circ$  for ACE, HMG, PAF, TXA2), giving diverse conformations
- One force field optimized conformation (Concord-generated) used to find other conformations of the same molecule in whole database of 937 structures, using Tanimoto Coefficient

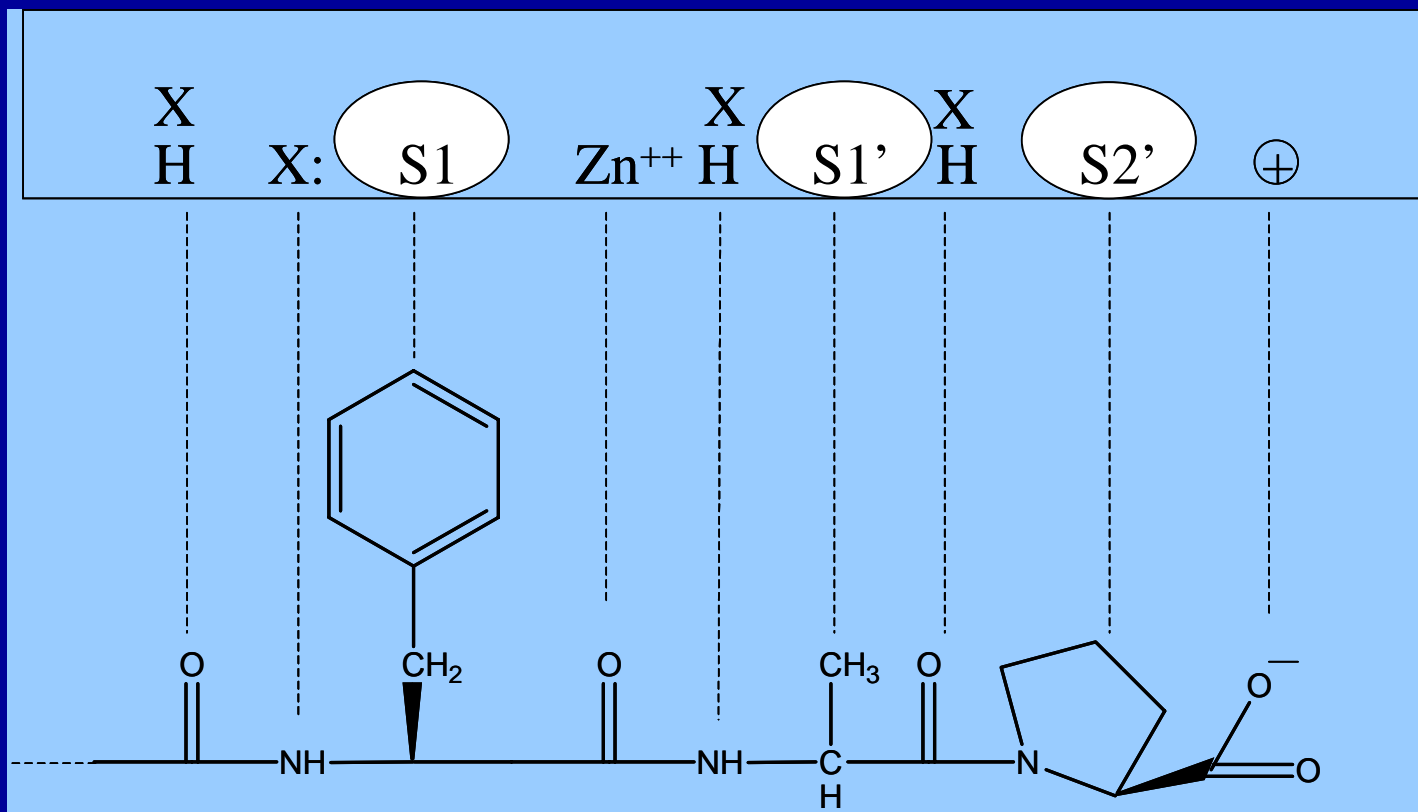
# Overall findings

- >90% of conformations found in Top 5% of sorted database
- Conclusion: If molecules with the right features are present in the database, they will not be missed (in most cases) because they are represented by a particular conformation

# Which features are selected for classification?

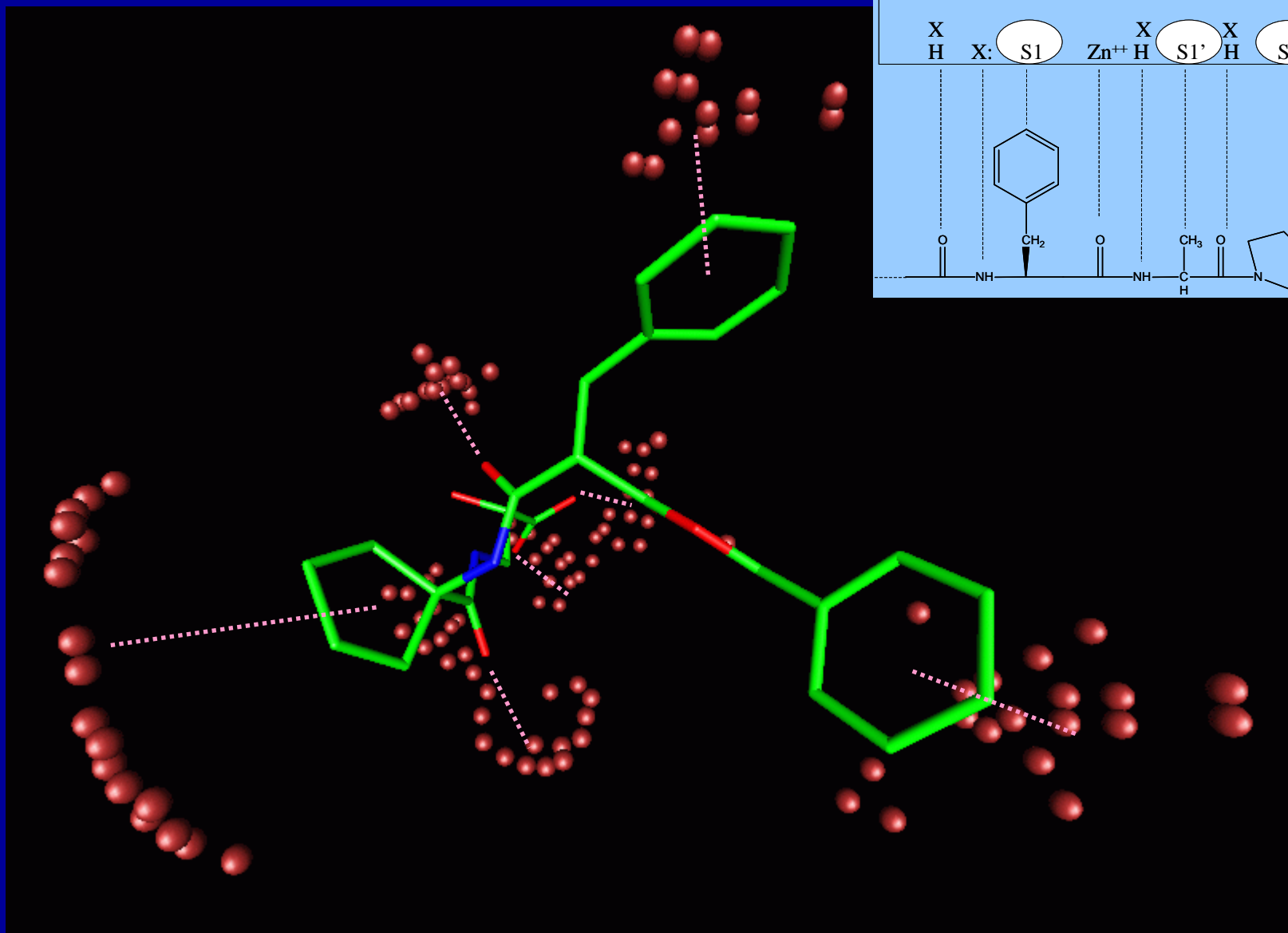
- Even if your classifier works, do the selected features make *sense*?
- Information Gain calculated for each feature, those which are much more frequent among actives are “suspicious” and might constitute a pharmacophore
- Look at features from ACE and TXA2 as examples

# ACE – Binding Site

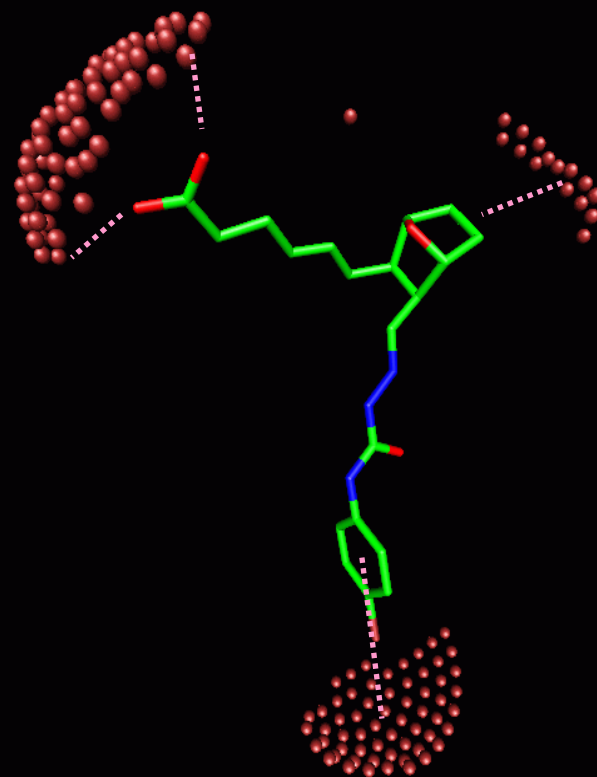
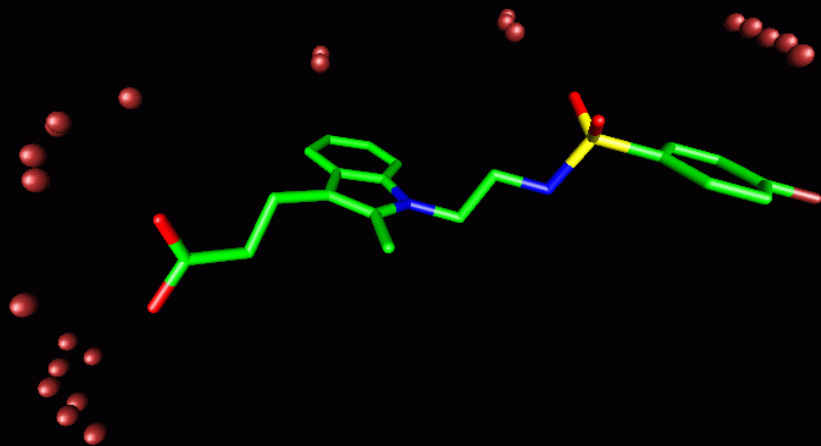
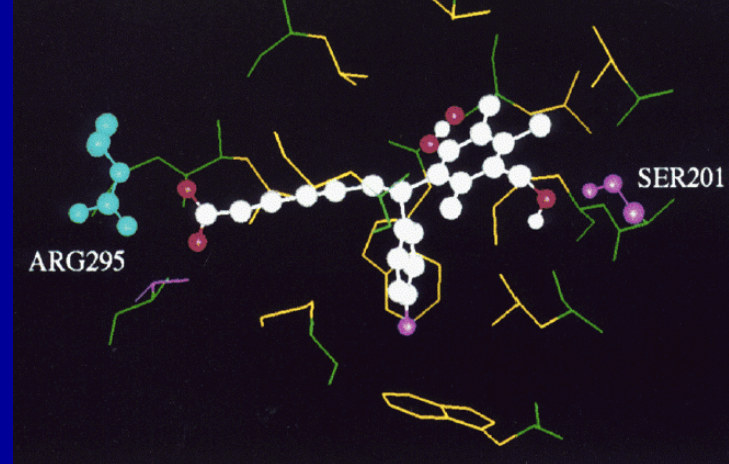


Snake venom peptide analog with putative binding motif to angiotensin used in early compound design (Cushman et al., *Biochemistry* (1977), 16, 5484-5491.) – recent crystal structure available

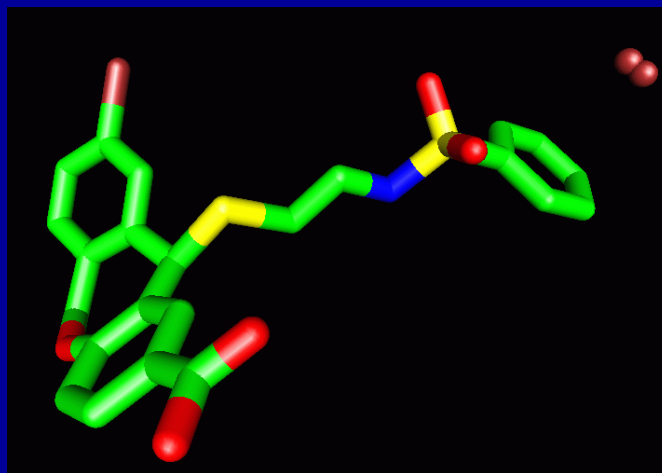
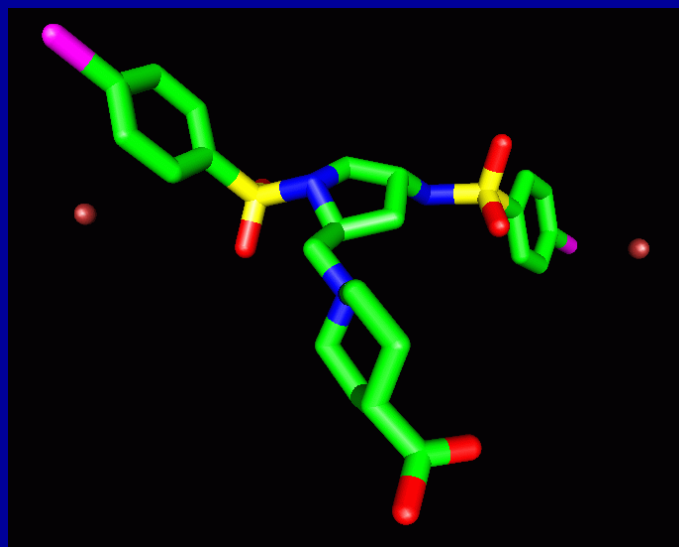
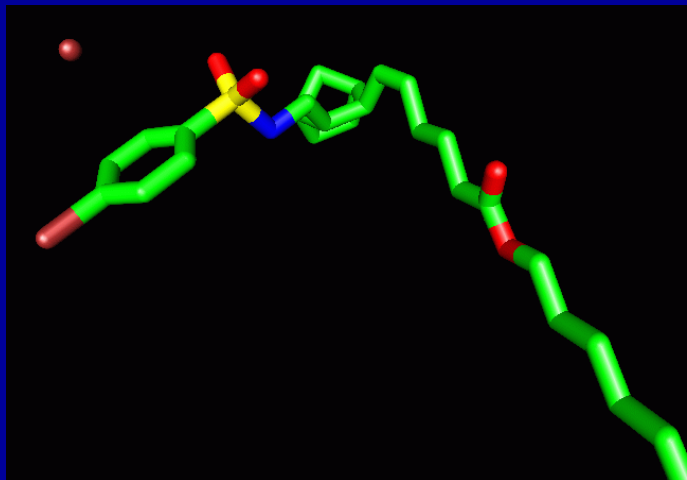
# Selected Features – ACE-31

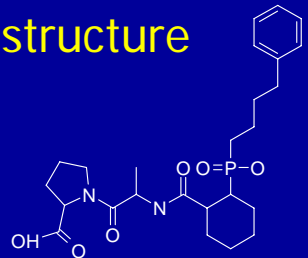







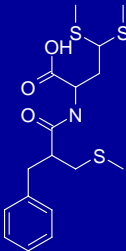
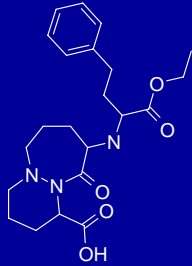
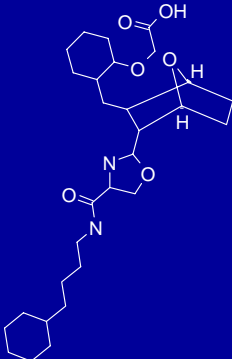
# TXA2- 7, and 44



# Most important feature of moving to 3D is "Structure Hopping"



Query	<p style="text-align: center; color: yellow; font-weight: bold;">Query structure</p> 		
Ranking position 1	Structure	Ranking position 2	Structure
3		4	
5		6	

7		8	
9		10	

Query (ACE inhibitor) used to screen the database and the highest ranked structures found (out of which all except no. 6,7 and 10 are classified as being ACE inhibitors in the MDDR database). **Five of the active structures found (no. 3, 4, 5, 8 and 9) were not found by any of the other seven methods employed. Maybe they are active? ?**

Molecular Surface Point Environments for Virtual Screening and the Elucidation of Binding Patterns (MOLPRINT 3D). Bender, A.; Mussa, H. Y.; Gill, G. S.; Glen, R. C. J. Med. Chem. 2004, 47(26), 6569-6583.

## A real test...HTS Data Mining and Docking Competition 2005 at McMaster University (Ontario)

A competition to take ~50,000 dihydrofolate reductase inhibitors of known activity (Training Set) and to (blindly) predict the activity of ~50,000 new compounds (Test Set) in a high throughput screen.

32 groups took part. We obtained what were ranked as some of the 'best' results.

MOLPRINT 2D, was employed using an original training set of 49,995 compounds, enrichment factors (between one and three) could be achieved on a test library, comprising 50,000 structures

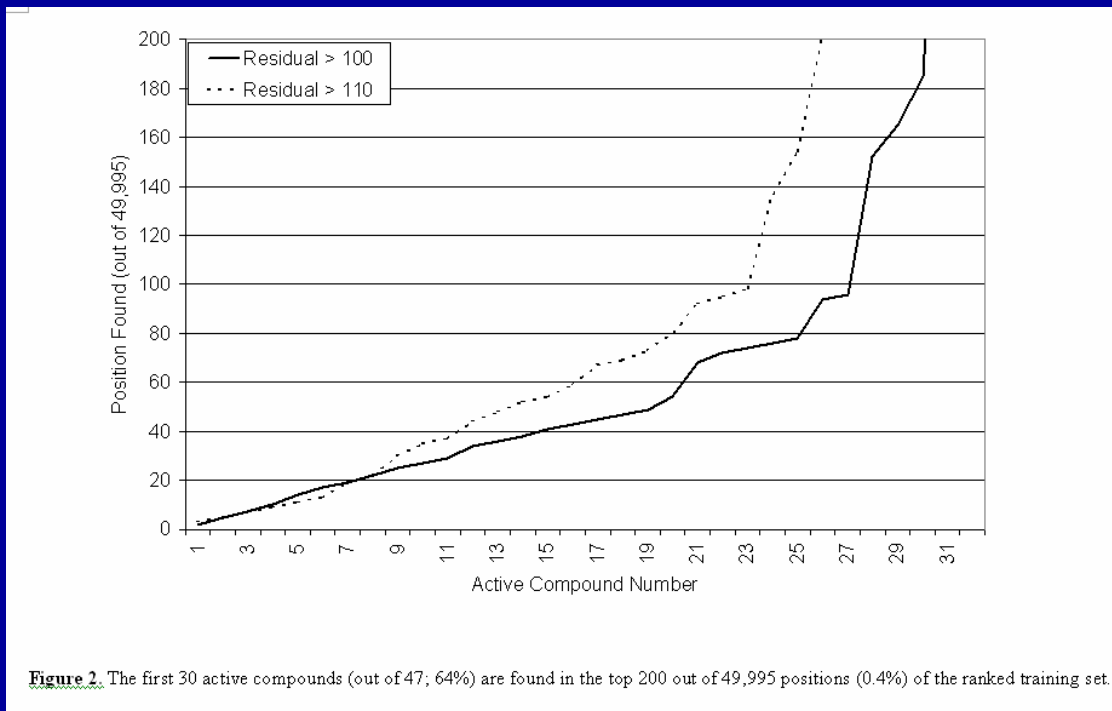
We think that these results are poor. Reasons are described below.

## Data Set :

High-throughput screening of 49,995 compounds was performed by Zolli-Juran *et al.*, identifying 32 hits (defined by less than 75% residual activity in both of two screening runs) comprising several novel scaffolds.

## Objective:

The extraction of the structural 'knowledge' from the compounds and their activities from the first screening ('training set') and to make predictions about the inhibitory activities of a second set of 50,000 compounds that was to be screened subsequently (42 weak 'hits' subsequently found in the 'test set').



Our results show ca. 3 fold enrichment in the first 200 compounds ranked. However, this reduced to just over one in the complete set – why ?

# Results

MolPrint2D

Number of Active Compounds Identified in Each Group's Ranked List

Group	# Submitted <sup>a</sup>	Consensus Residual Activity <sup>b</sup>		Average Residual Activity <sup>c</sup>		Comment <sup>e</sup>
		Active <sup>b</sup>	Well-Behaved <sup>d</sup>	Active <sup>c</sup>	Well-Behaved <sup>d</sup>	
1	50000	4	1	6	2	---
2	495	0	0	0	0	---
3	22	0	0	0	0	---
4	50	0	0	1	0	---
5	2000	0	0	3	0	---
6	127	0	0	0	0	---
7	50000	2	0	7	2	---
8	150	0	0	0	0	---
9	20	0	0	0	0	---
10	200	0	0	0	0	---
11	30	0	0	0	0	---
12	77	0	0	0	0	---
13	59	0	0	0	0	---
14	294	0	0	0	0	---
15	46901	1	1	4	1	---
16	344	0	0	0	0	---
17	10	0	0	0	0	---
18	21	0	0	0	0	---
19	105	0	0	0	0	---
20	50000	1	1	1	1	---
21	59	0	0	0	0	---
22	44	0	0	0	0	YES
23	6	1	0	1	0	YES
24	40	0	0	0	0	---
25	28	0	0	0	0	---
26	21	0	0	0	0	---
27	121	0	0	0	0	---
28	601	1	0	2	0	---
29	46720	2	2	13	5	YES
30	439	0	0	1	1	---
31	26	0	0	0	0	---
32	1000	0	0	0	0	---

a Total number of ranked compounds submitted by each group

*NOTE: If group submitted a larger list, only the top 2500 ranked compounds were used*

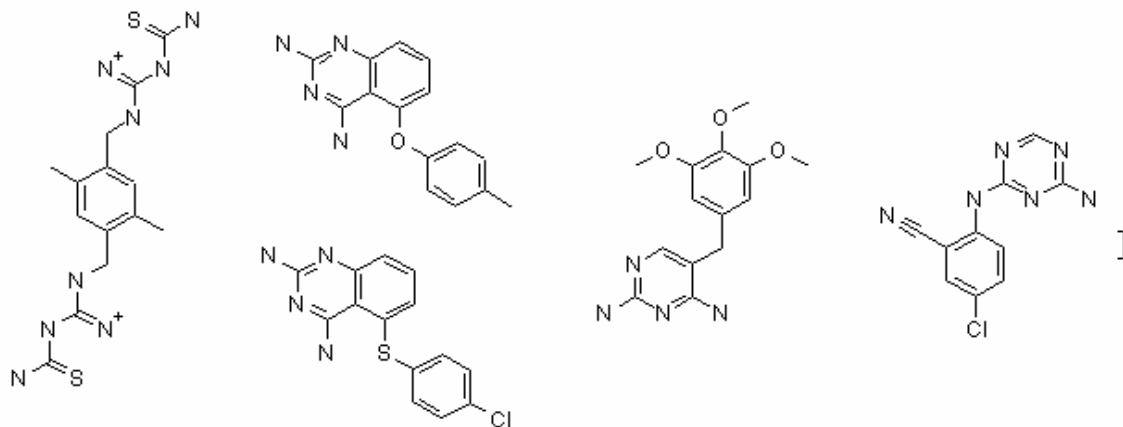
b Cutoff set at 75% residual activity for both replicates

c Cutoff set at 75% residual activity for the average of the replicates

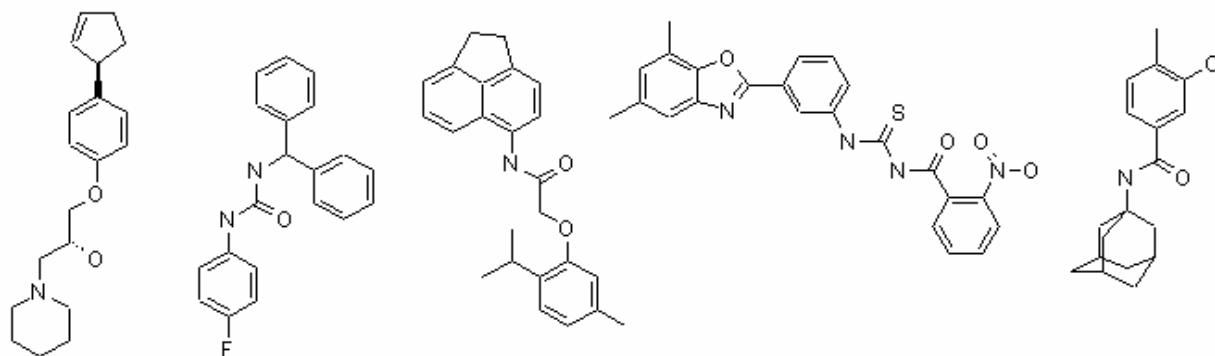
d Subset of active compounds for which a dose response curve could be obtained

e General comment made by group indicating knowledge either that the test set and the training set were from different areas of chemical space or that the test set would perform worse in this assay

Examples from  
the training set













Examples from  
the test set



Comparison of the most potent inhibitors of the training set (upper half) and the most potent inhibitors of the test set (lower half). Structural differences can already be identified on this small subset of compounds, for example the high number of pyridazine rings and guanidinium groups in the training set.

Features showing highest information-gain in discriminating active structures of the training set from those of the test set. Features characteristic for the most active compounds of each set are also more frequent in the whole set; this ratio is even more apparent among the active structures of each set.

Characteristic Features of Training-Set Active Structures				
	Number in Training Set	Number in Test Set	Among Actives in Training Set	Among Actives in Test Set
	416	159	10	0
	295	72	10	0
	40	72	8	0
	106	12	10	0
	136	0	9	0
Characteristic Features of Test-Set Active Structures				
	9077	18645	14	133
	6580	19186	4	126
	2449	10685	0	76
	9191	16043	12	115
	15202	25851	22	160

The 'Test Set' and the 'Training set' contains chemically different structures.

Therefore, this method does not always recognise new features in the new set as contributors to activity since it is based on the molecular graph.

We repeated the analysis by randomizing ALL the data and predicting using MolPrint2D and with cross validation to check for robustness

(standard QSAR post-hoc rationalisation !)

Results of training and test set after pooling in a second step and randomly splitting into training and test of equal size again, thus smoothing out the different chemical characteristics of both libraries.

	Hit Rates			Enrichment Factors		
First ... positions	96	384	1536	96	384	1536
Actives < 80% activity; Inactives > 100% activity, 200 Features	2	4	10	3.4	1.7	1.1
Ten-fold Random Validation Actives < 85% activity, Inactives > 100% activity, 200 Features	6.0 (0.7)	10.2 (2.4)	28.0 (3.0)	10.2 (1.2)	4.2 (1.0)	3.0 (0.3)

'Blind study'

after randomization - note  
big increase in success

In a ten-fold cross validation study on the new training and test sets, typically 10-fold enrichment could be found in the first 96 positions, 4-fold enrichment in the first 384 positions and 3-fold enrichment in the first 1536 positions, corresponding to 6, 10 and 28 hits (out of a total of 307), respectively.

Conclusions :

On the one hand the work presented here shows that exact-fragment-matching similarity searching methods are not capable of finding completely novel hit structures. Still, they are able to combine knowledge from multiple active structures to give novel combinations of features, as shown previously. On the other hand this work emphasizes the need for an even distribution of "chemistry" between the training and the test set. 'Lead hopping', moving from one chemical space to another thus requires analysis based on chemical descriptors (not the structural diagram), which is generally a much more compute intensive calculation.

# Summary

- 2D Method: Performs about as well as other 2D methods for single molecule searches, outperforms them by a large margin when combining information from multiple molecules
- 3D Method: TR invariant, conformationally tolerant; combines high enrichment factors with scaffold hopping – discovery of new chemotypes
- Features shown to correlate with binding patterns
- Performance (at least in part) due to Bayesian Classifier, which is able to take multiple structures as well as active *and* inactive information into account
- Chemically similar training and test sets required for 2D method
  - Bender A, Mussa HY and Glen RC. Screening for DHFR inhibitors using MOLPRINT 2D, a fast fragment-based method employing the Naïve Bayesian Classifier: Limitations of the descriptor and the importance of balanced chemistry in training and test sets. J Biomol Screen.2005; 10: 658-666

# However, The King has no clothes.....

We have also performed virtual screening using some **very simple features**; by employing the number of atoms per element as molecular descriptors, but without regard to any structural information whatsoever. Surprisingly (at least to me), these atom counts are able to outperform 'sophisticated' fingerprint approaches in some activity classes.

For all compounds of each dataset, simple atom counts were calculated : namely the total number of atoms, the number of heavy atoms and the numbers of Boron, Bromine, Carbon, Chlorine, Fluorine, Iodine, Nitrogen, Oxygen, Phosphorus and Sulfur atoms. Thus no structural descriptors at all were contained in this "fingerprint" representation which, besides the compound ID, contains just 12 integer numbers describing the frequency of different elements in the molecule.

The first dataset was published by Briem and Lessel<sup>6</sup> and it contains 957 ligands extracted from the MDDR database. The set contains 49 5HT<sub>3</sub> Receptor antagonists (5HT<sub>3</sub>), 40 Angiotensin Converting Enzyme inhibitors (ACE), 111 3-Hydroxy-3-Methyl-Glutaryl-Coenzyme A Reductase inhibitors (HMG), 134 Platelet Activating Factor antagonists (PAF) and 49 Thromboxane A<sub>2</sub> antagonists (TXA<sub>2</sub>). An additional 574 compounds were selected randomly which did not belong to any of these activity classes. The second and larger dataset was presented recently by Hert et al. 11 sets of active structures were defined, ranging in size from 349 to 1236 structures.

# There is previous Work on 'dumb' descriptors...

- Livingstone<sup>1</sup>: “Overall molecular parameters which are able to discriminate between compounds showing different physicochemical or biological behavior. E.g., blood-brain barrier penetration is closely related to logP, and **electron density on a nitrogen atom in the HOMO of a set of aniline mustards and tumor inhibition can be related in a simple linear fashion.** “
- Pan<sup>2</sup>: “**Heavier molecules are favored by docking algorithms** due to the simple fact that on average more atom-atom interactions are present which contribute to the predicted binding energy. As a remedy normalization of the binding energy with respect to the number of heavy atoms per molecule was suggested.”

<sup>1</sup> Livingstone, D. J. The characterization of chemical structures using molecular properties. A survey. *J. Chem. Inf. Comput. Sci.* 2000, 40, 195-209.

<sup>2</sup> Pan, Y. P., *et al.*, Consideration of molecular weight during compound selection in virtual target-based database screening. *J. Chem. Inf. Comput. Sci.* 2003, 43, 267-272.

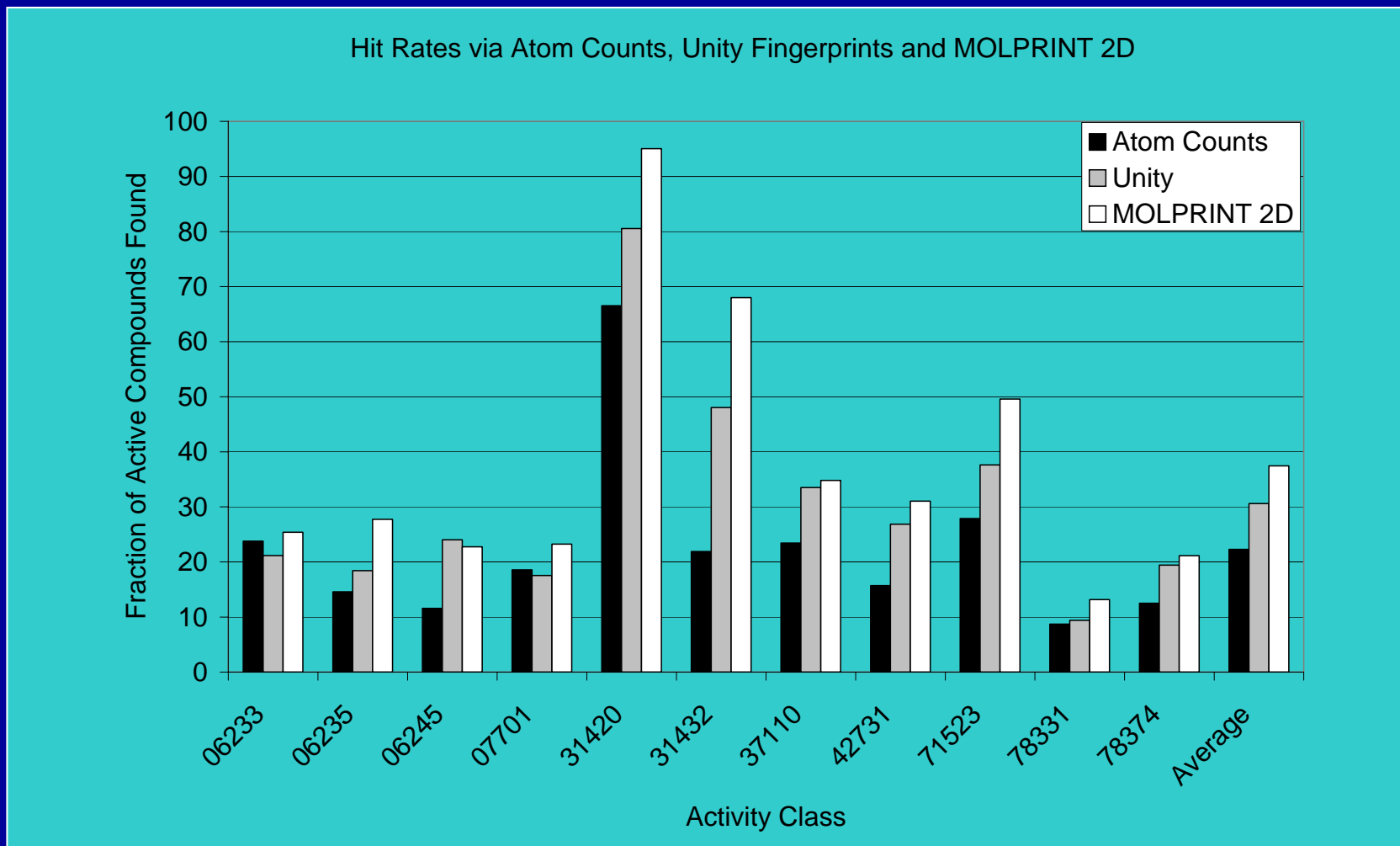
# Previous Work...

- Gillet<sup>3</sup>: “Bioactivity profiles (BPs) include the number of H-bond donors and acceptors, MW, a kappa shape index and the numbers of rotatable bonds and aromatic rings. BPs found application in distinguishing molecules from the World Drug Index and those from the SPRESI database (which were assumed to be inactive); **using single features such as the number of H-bond donors alone enrichments of up to 4.6** were found in identifying WDI molecules in a merged dataset. “
- Verdonk<sup>4</sup>: “Considering **heavy atom counts alone** on two hypothetical libraries of active compounds, which are either on average much heavier or much lighter than the whole library, was shown to give considerable enrichments. “

<sup>3</sup> Gillet, V. J.; Willett, P.; Bradshaw, J. Identification of biological activity profiles using substructural analysis and genetic algorithms. *J. Chem. Inf. Comput. Sci.* 1998, 38, 165-179.

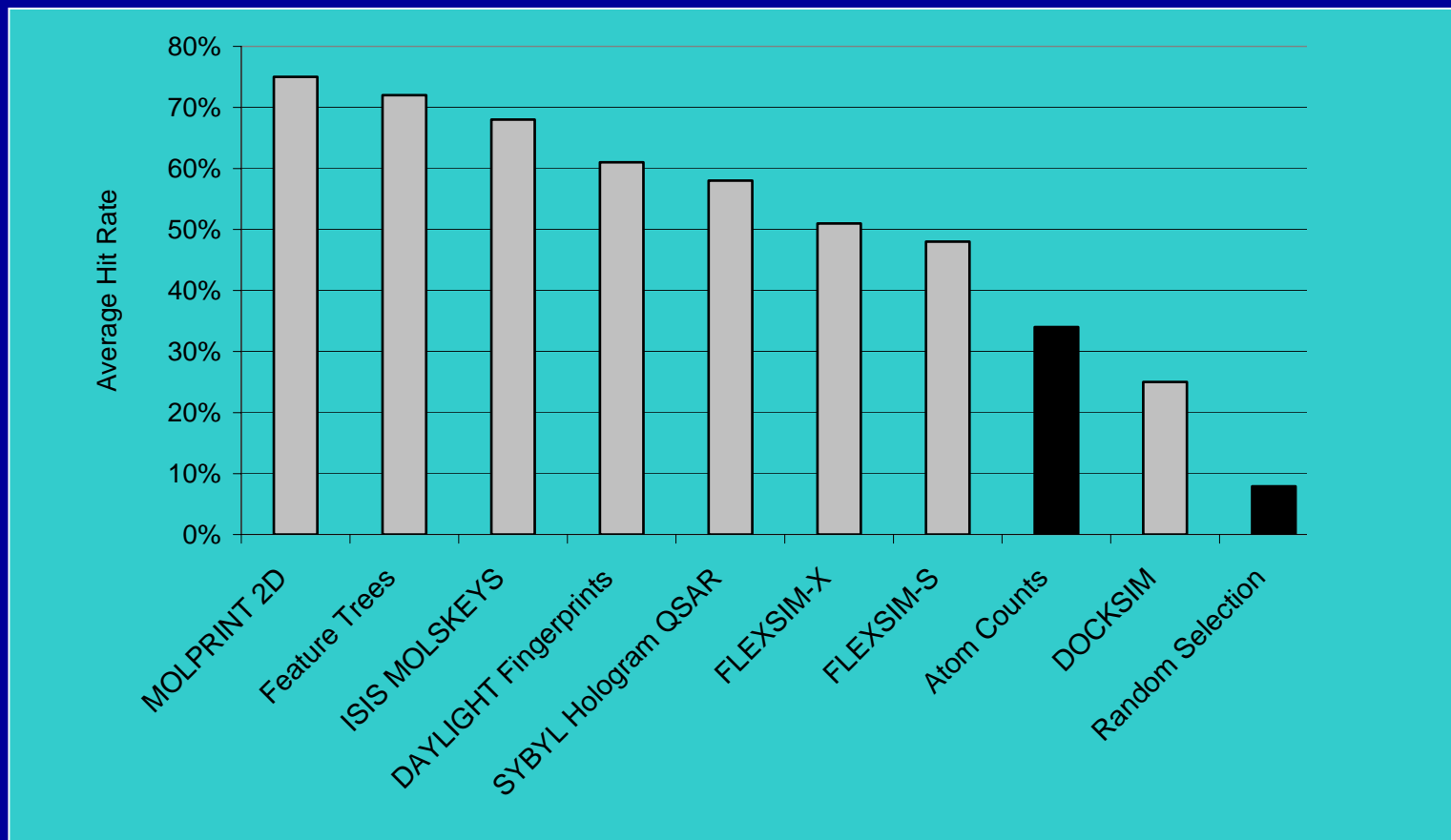
<sup>4</sup> Verdonk, M. L., *et al.*, Virtual screening using protein-ligand docking: Avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* 2004, 44, 793-806.

# Comparing retrieval rates for different targets



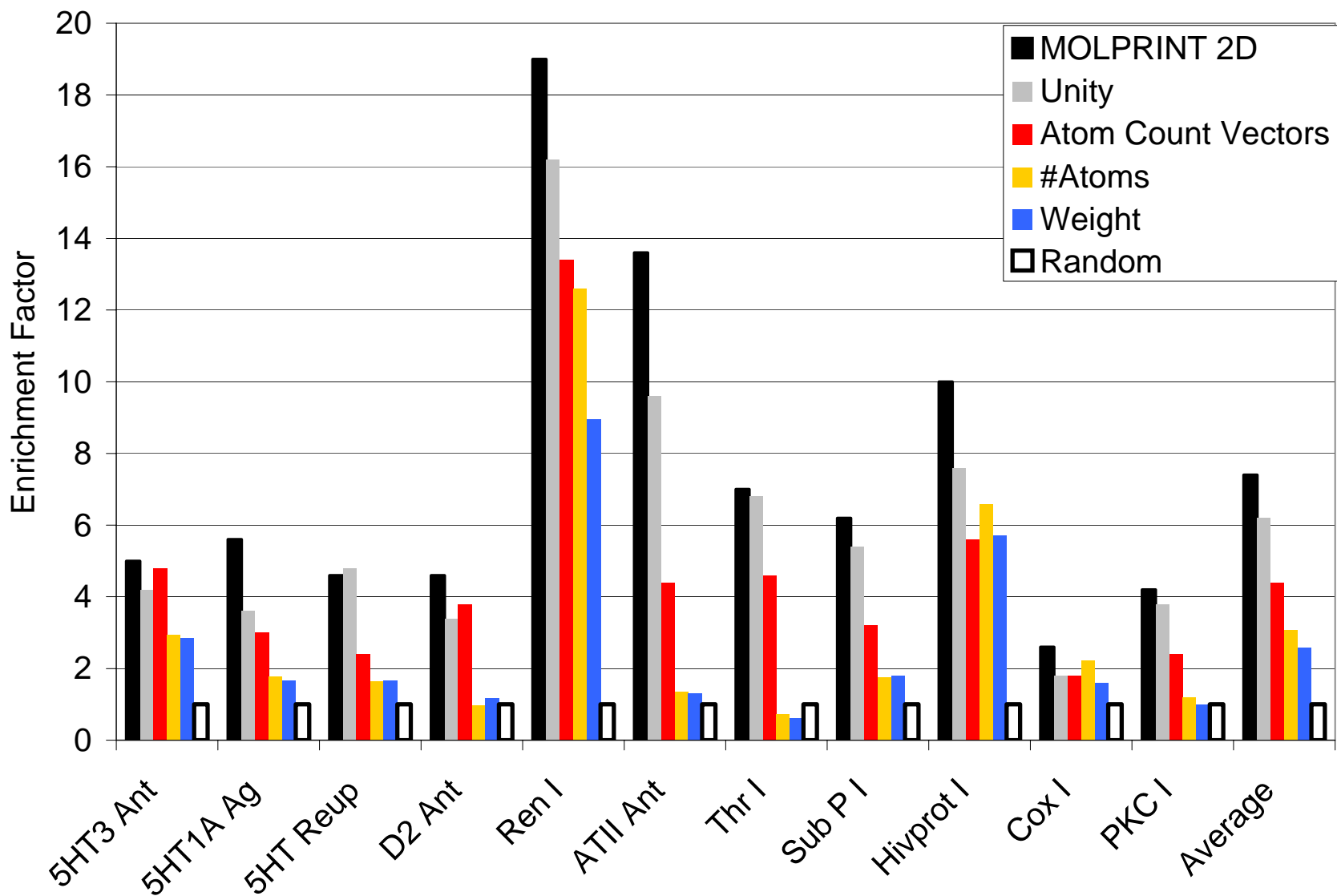
Fraction of active compounds found using simple atom counts, in comparison to Unity/MolPrint2D fingerprints. While Unity fingerprints outperform atom counts overall this margin is smaller than one might expect, given the fact that atom counts do not contain any structural information whatsoever while e.g. Unity fingerprints have some of that information available.

## Comparing retrieval rates for different methods



The average hit rate using “dumb” atom count-descriptors, compared to a variety of 2D and 3D similarity searching methods. Atom count descriptors achieve an enrichment of about 4-fold which is already superior to one of the virtual affinity fingerprint methods, DOCKSIM and around half the enrichment achieved by other methods employed!

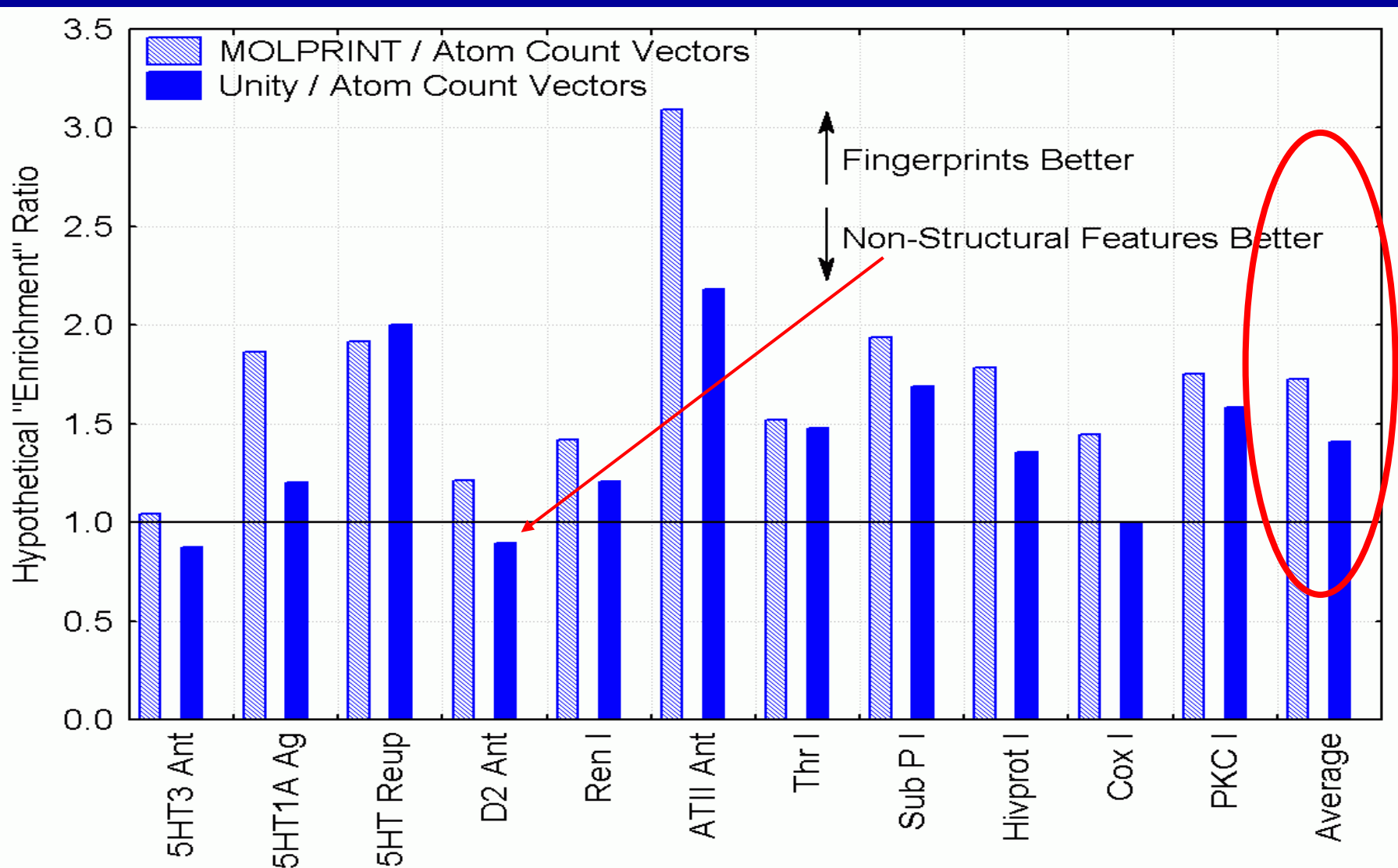
Interestingly, the Molecular Weight /#Atoms is not as good (compare red and blue)



Using simple atom count descriptors, up to more than ten-fold enrichment can be observed which is close to results achieved using Unity fingerprints/Tanimoto on the same dataset. So...we could try **Dividing the enrichment of the 'sophisticated' method by the 'dumb' method to better weight the usefulness of the methodology – this gives enrichment rates much closer to those observed in a 'real' experiment**

Activity Class	0623 3	0623 5	0624 5	0770 1	3142 0	3143 2	3711 0	4273 1	7152 3	7833 1	7837 4	Average
Hit Rate Atom Counts	23.78	14.59	11.59	18.58	66.53	21.89	23.42	15.69	27.89	8.69	12.48	22.28
Enrichment	4.76	2.92	2.32	3.72	13.31	4.38	4.68	3.14	5.58	1.74	2.50	4.46
Hit Rate Unity	21.15	18.43	24.02	17.53	80.54	48.04	33.51	26.87	37.60	9.39	19.42	30.59
Unity / Atom Counts	0.89	1.26	2.07	0.94	1.21	2.19	1.43	1.71	1.35	1.08	1.56	1.43
Hit Rate MOLPRINT 2D	25.40	27.73	22.75	23.24	95.04	68.01	34.79	31.03	49.56	13.16	21.13	37.44
MOLPRINT 2D / Atom Counts	1.07	1.90	1.96	1.25	1.43	3.11	1.49	1.98	1.78	1.51	1.69	1.68

# Dividing the enrichment of the 'sophisticated' method by the 'dumb' method



## Conclusion (what I think) when using structural motifs in virtual screening:

Databases of molecules are not random collections of molecules. They only contain a tiny fraction of possible molecules – and most of them are rather similar (maybe not to the receptor, but in terms of chemical fragments). Seeding a database with 'actives' allows an algorithm to 'induce' clear features for recognition – actually often quite simple features. Finding the actives again from the database is simple – they've been memorised – differentiated by simple features. So how 'good' is a new method ?

Simple atom counts can select activity classes. A better measure of success of a new screening method compared to 'random' selection would be to divide the results using a 'banal' feature like atom counts. This would give a better 'real' measure of the performance of 'sophisticated' methods when dealing with the problem of bias in molecule datasets.

# Acknowledgements

- Peter Murray-Rust, Jonathan Goodman, Hamse Mussa, **Andreas Bender**, Joe Townsend, Yong Zhang, Simon Tyrrell, Scott Boyer, James Smith, Catrin Hasselgren Arnby, Lars Carlsson, Li Xing, Bob Clark
- Unilever, the Royal Society of Chemistry, the Newton Trust, the Department of Trade and Industry, the EPSRC, the BBSRC.
- And thanks to the UK QSAR community