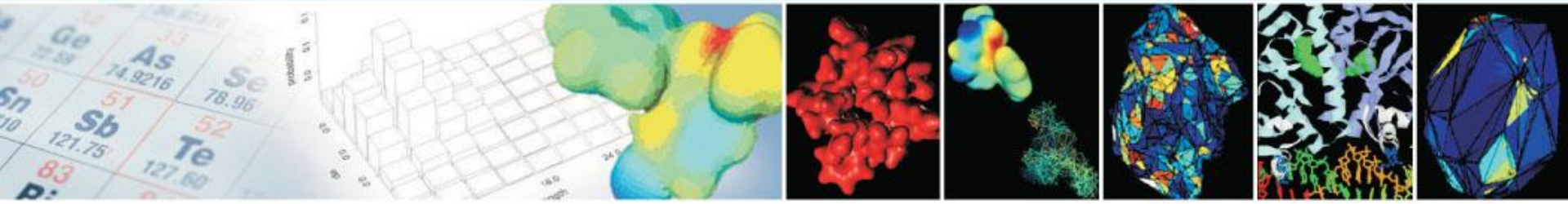
RECCR
Rensselaer Exploratory Center for Cheminformatics Research



Predictive Cheminformatics Strategies for Anticipating Good and Bad Side Effects -

New Methods for predicting Multiple CYP Metabolic Sites and Off-target Polypharmacology

Curt M. Breneman*, Kristin P. Bennett, Jed Zaretzki, Mark Embrechts,
Charles Bergeron and Sourav Das

Columbia University and Schrodinger, Inc. Conference on Computer-Aided Drug Design
June 18, 2010

Presentation Outline

- Part I. Metabolic regioselectivity models for nine CYP450s using RS_Predictor (Jed Zaretzki, Charles Bergeron and Kristin Bennett)

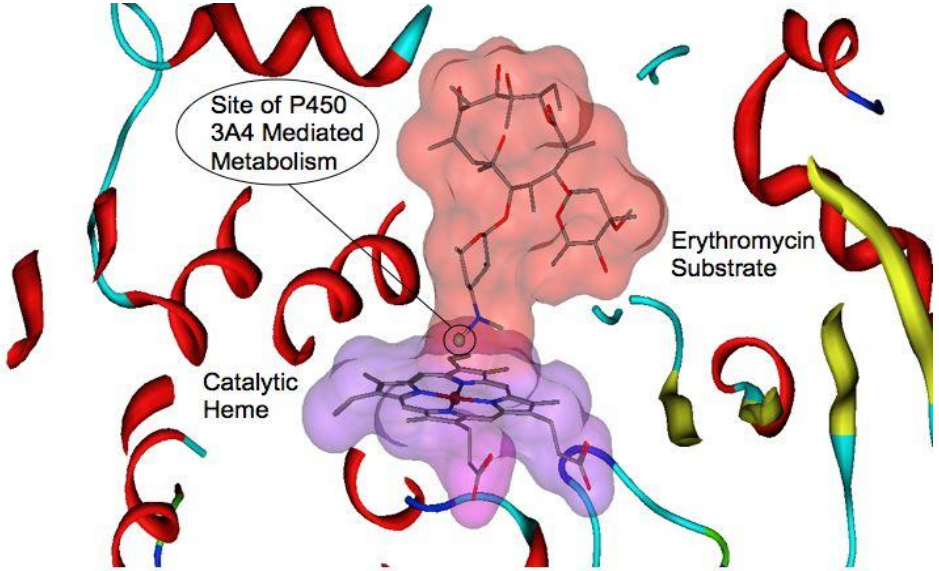
- Part II. Property-Encoded Shape Distributions (PESD) for Comparing Protein Binding Sites and Predicting Off-target Interactions (Sourav Das)

Part I. Metabolic regioselectivity models for nine CYP450 isozymes using RS_Predictor

(Jed Zaretzki, Charles Bergeron and Kristin Bennett)

Overview of Part I

- Motivation
- Identify the problem
- Methods
- Datasets
- Results
- Conclusions



Motivation: Why is this important?

- Cytochrome P450s account for approximately 90% of phase I metabolic reactions of all marketed drugs
- Prediction of metabolic sites on lead candidates empowers medicinal chemists to:
  - modify labile sites of lead candidates in order to increase bioavailability without changing efficacy
  - perform pro-drug design
  - Identify and block potential metabolites with undesired PK behavior
- Reliable *in silico* identification of metabolic liabilities early in the drug discovery process would allow early triage or modification of unsuitable lead compounds

Motivation: What's come before

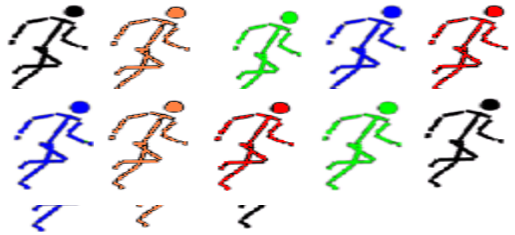
- **Reactivity-Based Models – ligand only**
  - QSAR-based regioselectivity models using a random forest algorithm (Sheridan et al., 2007)
  - AM1 Semi-empirical calculations (Singh et al., 2003) used to estimate the energy necessary to abstract a hydrogen atom from a substrate
- **Recognition-Based Models – ligand and enzymatic structure**
  - MetaSite reactivity and recognition-based application (Cruciani et al., 2005) utilizing GRID molecular interaction fields (Goodford et al., 1985)
  - Docking algorithms, Dock (Ewing et al., 2001), Glide (Friesner et al., 2004), and GLUE (Zamora et al., 2006)

Identifying the Problem:

A racing metaphor

Race 1

Race 4
Race 2

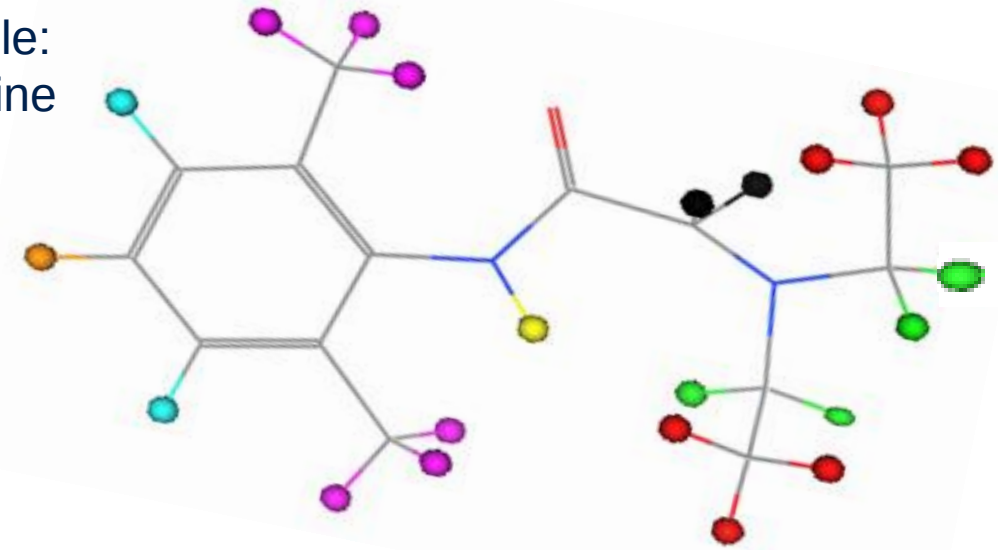


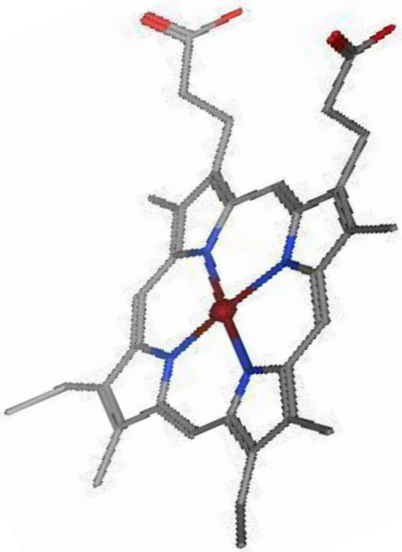
Race 3

FINISH

OXIDIZE

Example:
Lidocaine

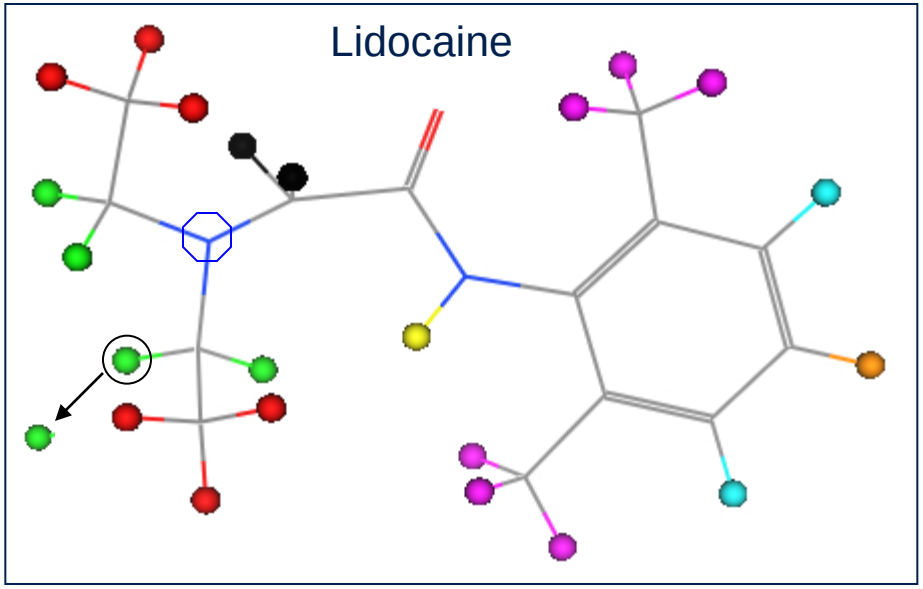


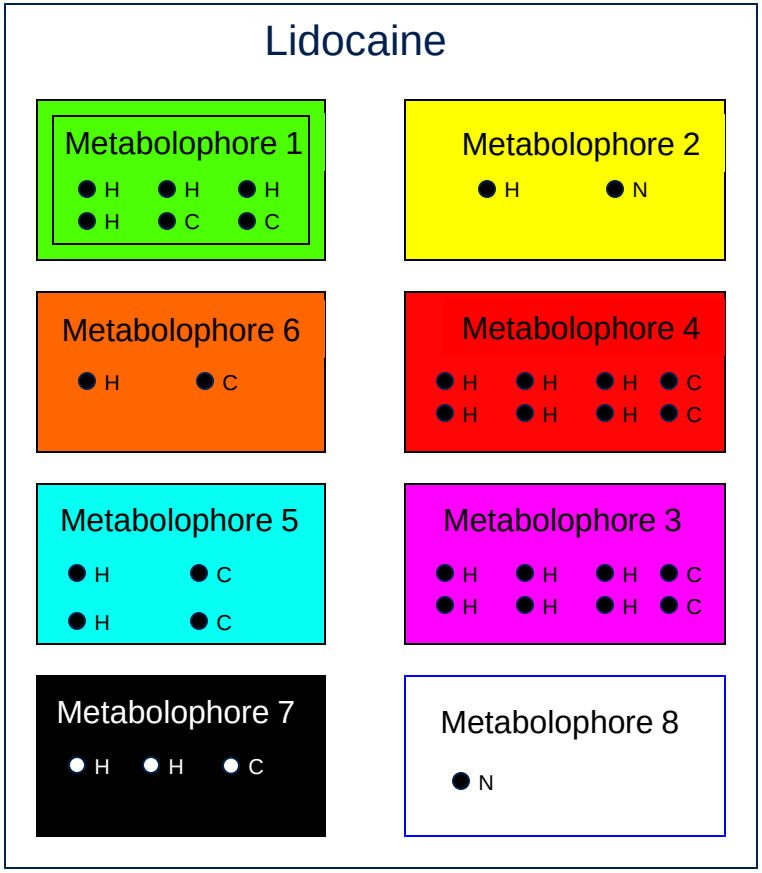


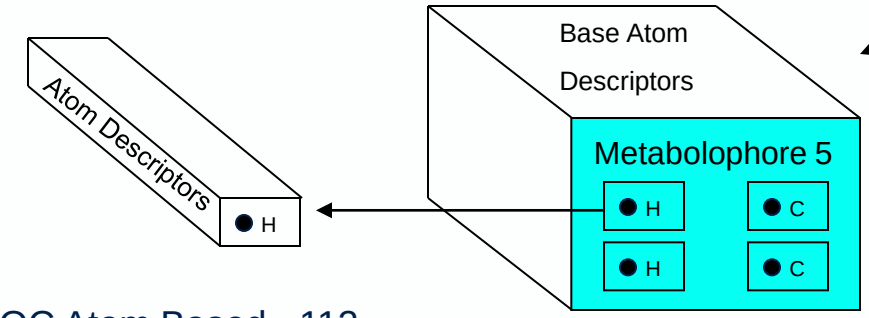
New Methods

- **RS-Predictor -** A specialized QSAR using Multiple-Instance Ranking (MIRank) and hierarchical electronic descriptors

- **SMARTCyp -** A 2D method using DFT transition state calculations on molecular fragments to create energy rules representing site reactivity







QC Atom Based - 112

- AM1 charge
- Hydrophobic moment
- Fukui reactivity

QC Atom Pair Based - 280

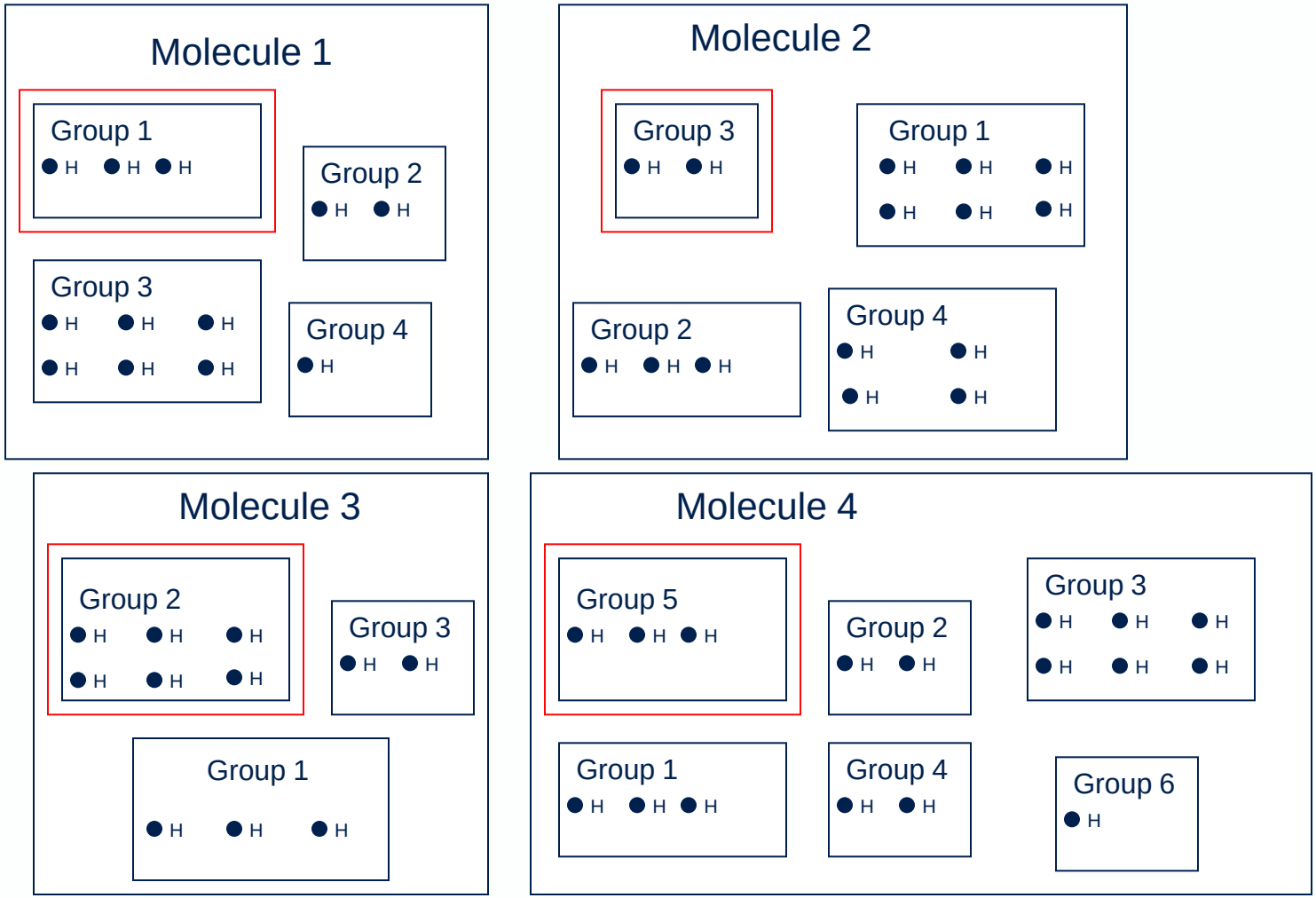
- σ − σ bond order
- Electronic resonance
- Coulomb interaction

Topological Descriptors - 148

- Hydrogen bond count
- Span
- Ring information
- Rotatable bonds
- Physical environment
  - •Distribution of atom types at 1, 2 , 3 and 4 bonds away from base atom

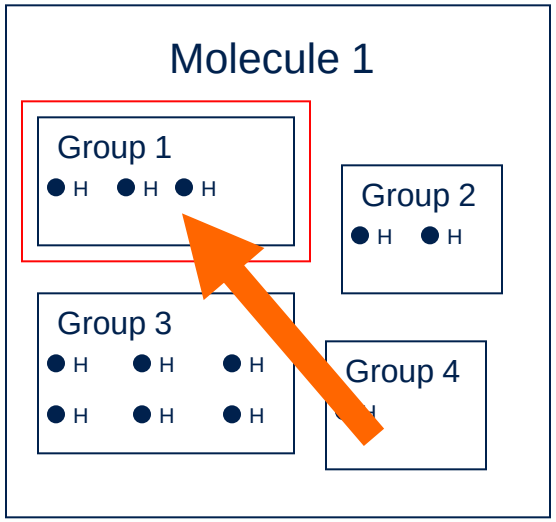
Green group designates the experimentally determined site of metabolism

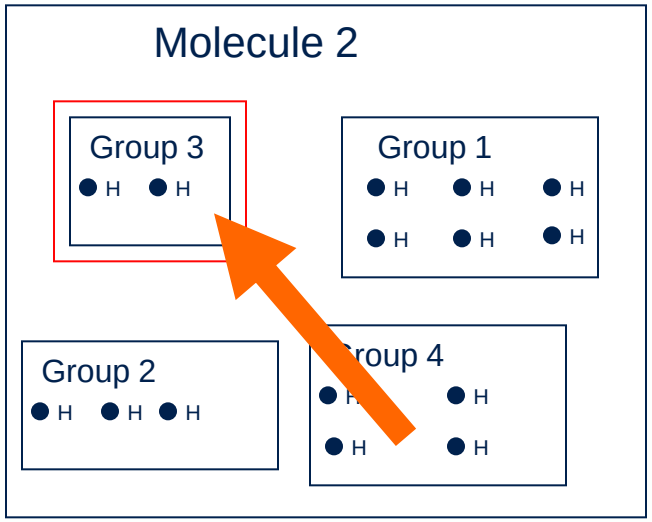
Trend Identification using Multiple-Instance Ranking (MIRank)

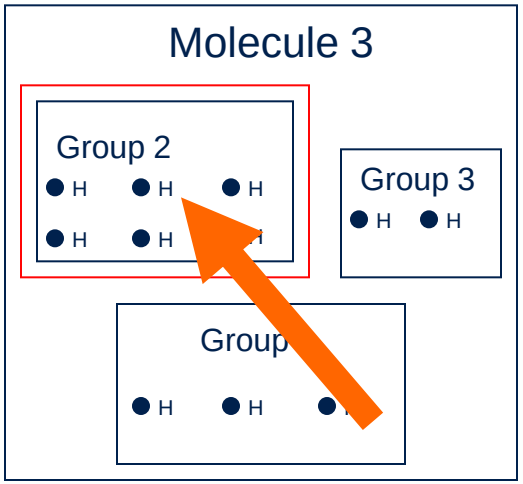


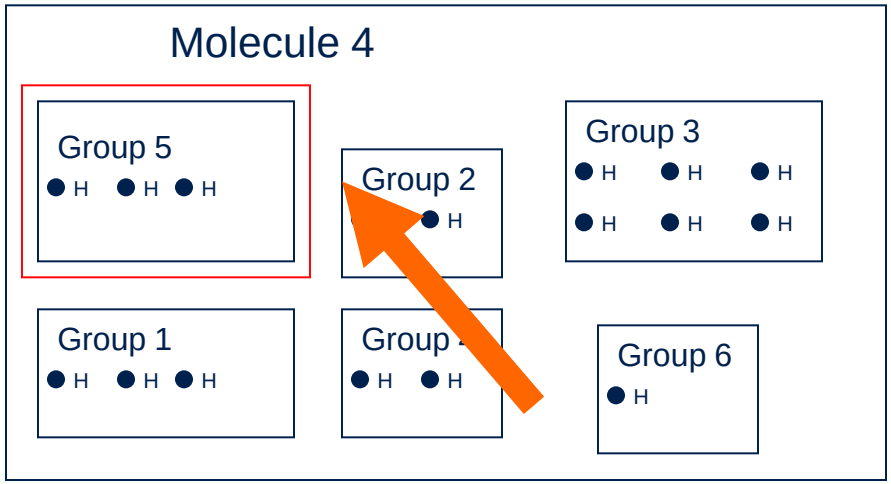
Multiple Instance Ranking

(Bergeron *et al.*, IEEE PAMI)









MIRank identifies descriptor-based trends present in each molecule

Trends are then combined to produce a single global ranking model of metabolic regioselectivity

Datasets

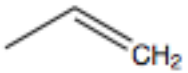
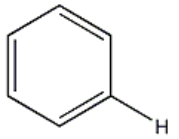
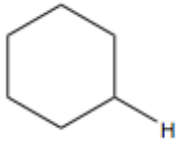
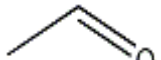
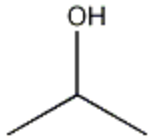
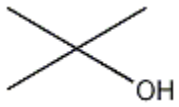
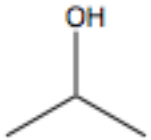
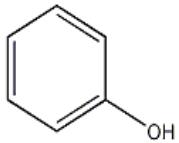
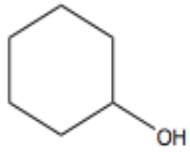
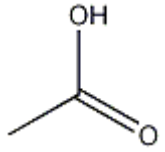
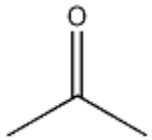
Prior to this work, few public datasets of P450 substrates with experimental responses existed - (Sheridan *et al.*, 2007)

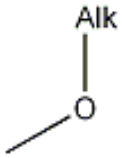
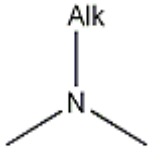
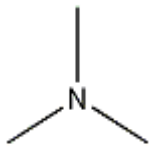
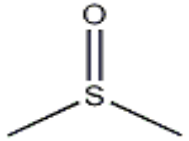
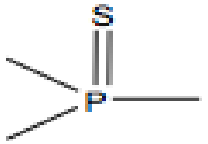
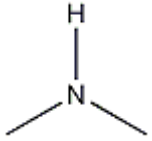
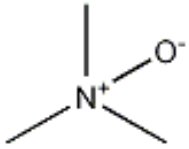
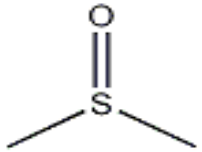
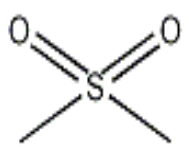
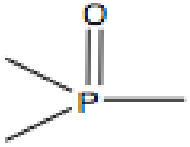
- 3A4 - 324 compounds
- 2D6 - 132 compounds
- 2C9 - 101 compounds

We have expanded these three datasets and created new datasets for nine isozymes:

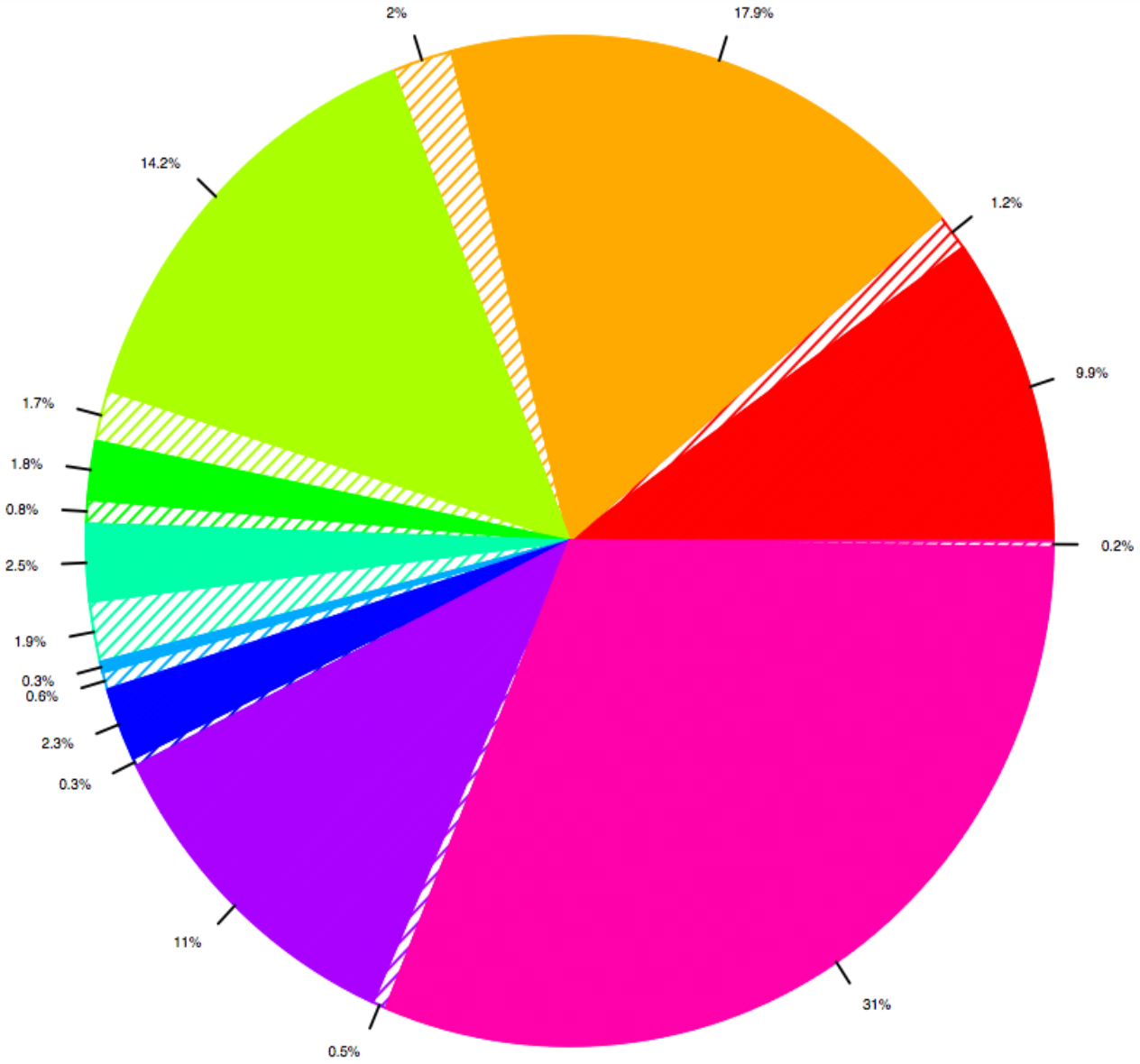
Isozyme	1A2	2A6	2B6	2C19	2C8	2C9	2D6	2E1	3A4
Size	256	97	127	192	120	209	256	117	459

Common CYP P450-mediated reactions

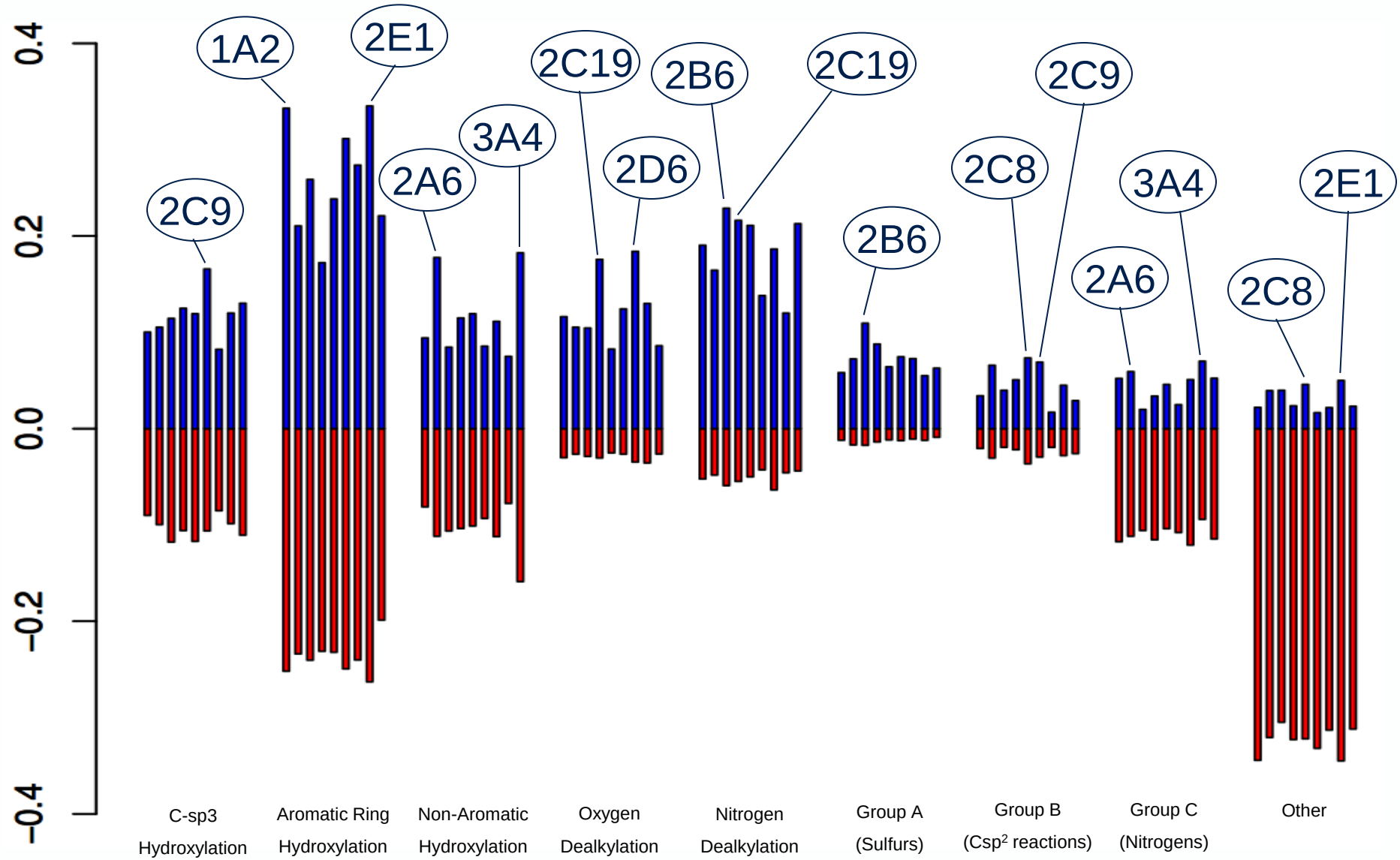
Reaction	C-sp3 Hydroxylation	C-sp2 Hydroxylation	Aromatic-Ring Hydroxylation	Non-Aromatic Ring Hydroxylation	Aldehyde Oxidation	Alcohol Oxidation
Initial Fragment						
Final Fragment						

Reaction	O-Dealkylation	N-Dealkylation	N-Oxide Formation	S(II) Oxidation	S(IV) Oxidation	Desulfuration
Initial Fragment						
Final Fragment						

Observed and Potential SOMs of 459 3A4 substrates broken down by reaction pathway

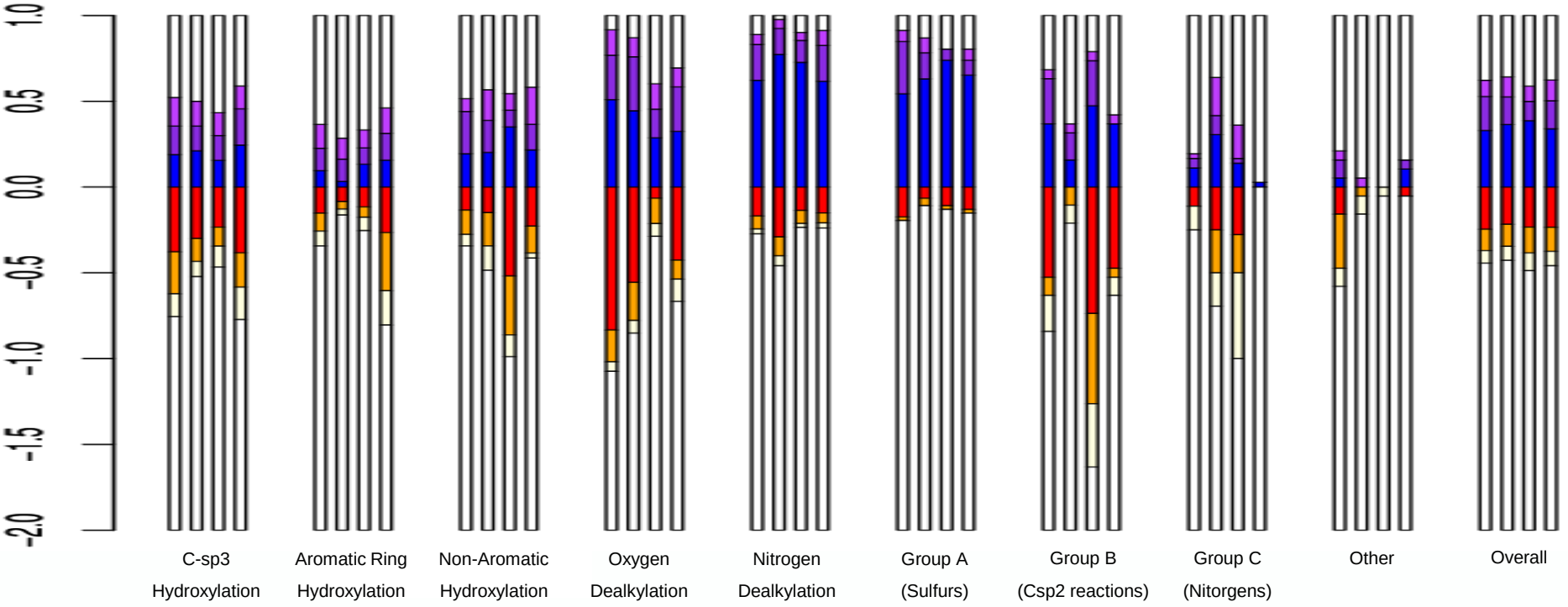


Pathway preferences (major column) by **Isozyme** (minor column)

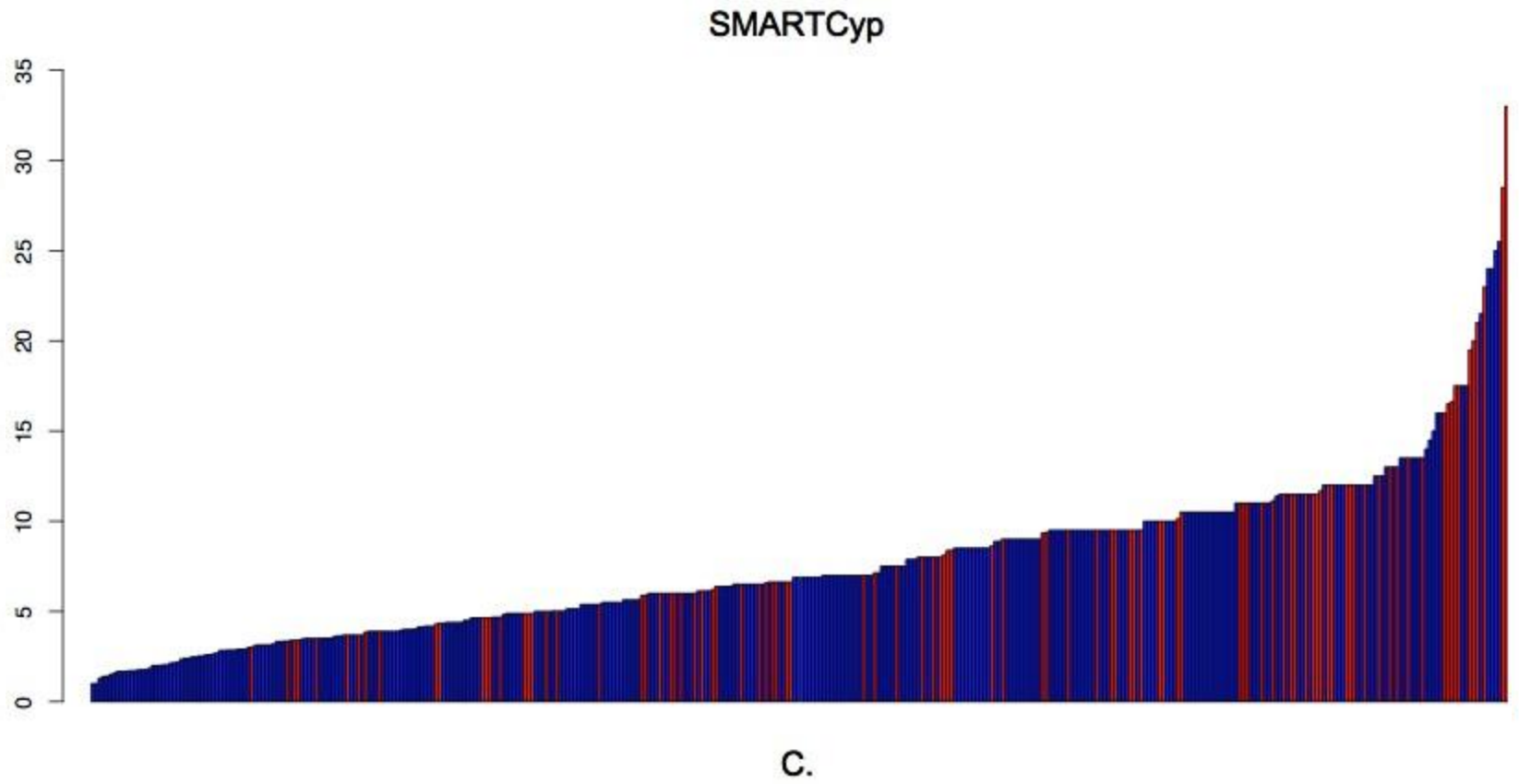


3A4 - Results (394 Compounds)

Method Metric	RS-Predictor	Metasite	SMARTCyp	Stardrop
Top 1	59.64%	62.5%	63.39%	58.76%
Top 2	79.70%	77.41%	73.16%	74.87%
Top 3	86.29%	85.55%	80.45%	83.76%



3A4 - Results (394 Compounds)



Number of potential sites of metabolism

Overall Results

Size	Isozyme	RS Top 1	RS Top 2	RS Top 3	SC Top 1	SC Top 2	SC Top 3
256	1A2	66.80%	81.25%	87.5%	62.01%	78.79%	87.66%
97	2A6	64.95%	79.38%	86.60%	62.71%	79.04%	90.23%
127	2B6	59.84%	74.02%	83.46%	60.63%	70.60%	83.27%
192	2C19	62.50%	77.08%	83.33%	55.90%	69.36%	79.98%
120	2C8	59.17%	75.00%	84.17%	56.28%	73.19%	83.44%
209	2C9	57.89%	74.16%	84.21%	56.53%	65.87%	78.95%
256	2D6	70.31%	82.81%	85.94%	42.77%	53.78%	64.31%
117	2E1	55.56%	74.36%	79.49%	57.69%	79.49%	82.91%
459	3A4	59.04%	77.56%	85.62%	62.03%	72.60%	81.34%

Cases where RS-Predictor outperforms SMARTCyp by > 5%

Size	Isozyme	RS Top 1	RS Top 2	RS Top 3	SC Top 1	SC Top 2	SC Top 3
256	1A2						
97	2A6						
127	2B6						
192	2C19	62.50%	77.08%		55.90%	69.36%	
120	2C8						
209	2C9		74.16%	84.21%		65.87%	78.95%
256	2D6	70.31%	82.81%	85.94%	42.77%	53.78%	64.31%
117	2E1						
459	3A4						

Cases where SMARTCyp outperforms RS-Predictor

Size	Isozyme	RS Top 1	RS Top 2	RS Top 3	SC Top 1	SC Top 2	SC Top 3
256	1A2			87.5%			87.66%
97	2A6			86.60%			90.23%
127	2B6	59.84%			60.63%		
192	2C19						
120	2C8						
209	2C9						
256	2D6						
117	2E1	55.56%	74.36%	79.49%	57.69%	79.49%	82.91%
459	3A4	59.04%			62.03%		

Current hypothesis is that SMARTCyp performs best on small molecules. 2E1 database contains a significant number of small compounds

What if we combined RS-Predictor and SMARTCyp?

Overall Results - SMART-RS-Predictor

Size	Isozyme	RSTOP + SC Top 1	RSTOP + SC Top 2	RSTOP + SC Top 3	RS + SC Top 1	RS + SC Top 2	RS + SC Top 3
256	1A2	65.63%	81.25%	90.23%	71.09%	84.77%	89.45%
97	2A6	70.10%	88.66%	91.75%	69.07%	82.47%	88.66%
127	2B6	66.14%	81.89%	88.19%	61.42%	80.31%	85.04%
192	2C19	65.63%	82.29%	88.54%	65.10%	81.77%	88.54%
120	2C8	63.33%	80.00%	89.17%	59.17%	81.67%	87.5%
209	2C9	64.59%	81.82%	87.08%	61.72%	77.99%	83.25%
256	2D6	69.92%	80.08%	87.12%	67.58%	83.59%	89.84%
117	2E1	63.25%	79.49%	85.47%	57.26%	73.50%	78.63%
459	3A4	65.14%	80.17%	87.15%	65.14%	80.17%	88.45%

Top 2 Metric - All methods

Size	Isozyme	RS	SC	RSTOP + SC	RS + SC
256	1A2	81.25%	78.79%	81.25%	**84.77%**
97	2A6	79.38%	79.04%	**88.66%**	82.47%
127	2B6	74.02%	70.60%	**81.89%**	80.31%
192	2C19	77.08%	69.36%	**82.29%**	81.77%
120	2C8	75.00%	73.19%	80.00%	**81.67%**
209	2C9	74.16%	65.87%	**81.82%**	77.99%
256	2D6	82.81%	53.78%	80.08%	**83.59%**
117	2E1	74.36%	**79.49%**	**79.49%**	73.50%
459	3A4	77.56%	72.60%	**80.17%**	**80.17%**
Average		77.29%	71.41%	**80.74%**	**80.69%**

Top 1 Metric - All methods

Size	Isozyme	RS	SC	RSTOP + SC	RS + SC
256	1A2	66.80%	62.01%	65.63%	**71.09%**
97	2A6	64.95%	62.71%	**70.10%**	69.07%
127	2B6	59.84%	60.63%	**66.14%**	61.42%
192	2C19	62.50%	55.90%	**65.63%**	65.10%
120	2C8	59.17%	56.28%	**63.33%**	59.17%
209	2C9	57.89%	56.53%	**64.59%**	61.72%
256	2D6	**70.31%**	42.77%	69.92%	67.58%
117	2E1	55.56%	57.69%	**63.25%**	57.26%
459	3A4	59.04%	62.03%	**65.14%**	**65.14%**
Average		61.78%	57.39%	**65.97%**	**64.17%**

Top 3 Metric - All methods

Size	Isozyme	RS	SC	RSTOP + SC	RS + SC
256	1A2	87.5%	87.66%	**90.23%**	89.45%
97	2A6	86.60%	90.23%	**91.75%**	88.66%
127	2B6	83.46%	83.27%	**88.19%**	85.04%
192	2C19	83.33%	79.98%	**88.54%**	**88.54%**
120	2C8	84.17%	83.44%	**89.17%**	87.5%
209	2C9	84.21%	78.95%	**87.08%**	83.25%
256	2D6	85.94%	64.31%	87.12%	**89.84%**
117	2E1	79.49%	82.91%	**85.47%**	78.63%
459	3A4	85.62%	81.34%	87.15%	**88.45%**
Average		84.48%	81.34%	**88.30%**	**86.60%**

Top 2 Metric - All methods

Size	Isozyme	RSTOP + SC	RS + SC	Merck	Metasite
101	2C9	**86.60%**	82.47%	77%	62%
132	2D6	84.00%	**86.26%**	62%	65%
324	3A4	**81.00%**	80.68%	63%	69%

Size	Isozyme	RSTOP + SC	RS + SC	Merck	Metasite
209	2C9	**81.82%**	77.99%		
256	2D6	80.08%	**83.59%**		
459	3A4	**80.17%**	**80.17%**		

Additional Results:
Private data

- Blind predictions were made on a set of 20 proprietary compounds provided by a partnering pharmaceutical company
    - Predictions were made using models developed from the public literature
- **RS-Predictor correctly predicted the observed sites of metabolism within the top two rankings in 85% of blind test compounds**
- **The correct region of metabolism was identified in 100% of the cases**

Summary of Part I. RS-Predictor

- RS_Predictor utilizes customized descriptors and exploits a novel machine learning framework to address a difficult problem with limited experimental data

- We have compiled and will release an extensive set of curated public P450 metabolic site data across nine isozymes, facilitating future research and applications

Future research

- Explicitly incorporate CYP binding site structural and electronic properties into prediction method.
- Apply multitask learning across CYPs by exploiting common catalytic trends present in all isozymes.
- Create isozyme substrate specificity models.

Graduating soon…

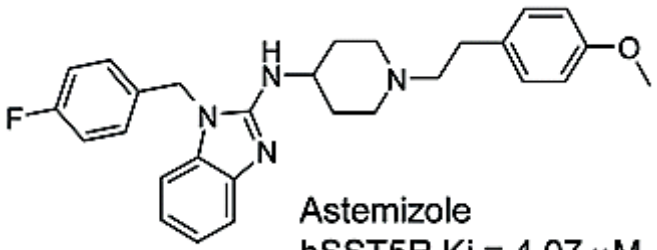
Part II. Property-Encoded Shape Distributions (PESD) for Comparing Protein Binding Sites and predicting Off-target Interactions

(Dr. Sourav Das)

Motivation

Discovery of low molecular weight somatostatin receptor subtype 5 (hSST5R) antagonists

- Astemizole as the lead structure
- Astemizole's original target was H1, a histamine receptor
- H1 has binding site amino-acid composition similar to hSST5R



Astemizole
hSST5R Ki = 4.07 μM

Martin, R. E.; Green, L. G.; Guba, W.; Kratochwil, N.; Christ, A. *J. Med. Chem.* **2007**, *50*, 6291-6294.

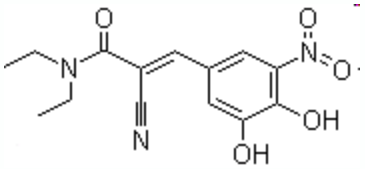
Motivation

Repositioning Entacapone : Discovery of safe chemical compounds with the potential to treat MDR-TB and XDR-TB

- Original target: Human catechol-*O*-methyltransferase (COMT)

- Binding site similarity: COMT and *M. tuberculosis* enoyl-acyl carrier protein reductase (InhA)

- Ligand docked and experimentally validated: activity MIC_{99} = 260 µM



Kinnings, S. L.; Liu, N.; Buchmeier, N.; Tonge, P. J.; Xie, L.; Bourne, P. E. *PLoS Comput. Biol.* **2009**, *5*, e1000423.

Motivation

Side-effects: Both good and bad

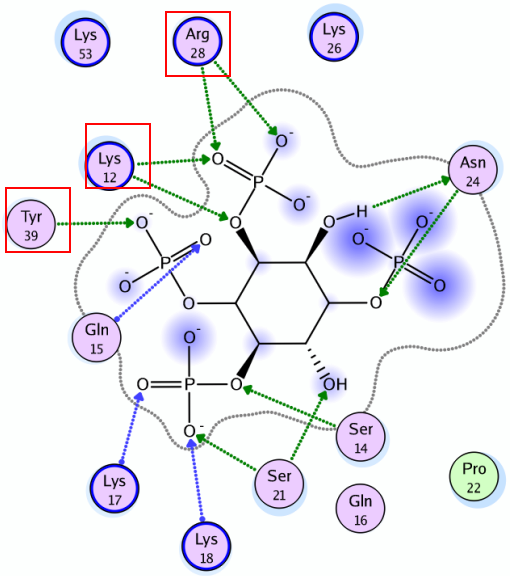
- Gleevec and Sutent act on several targets

- Permax and Dostinex activating 5-HT_{2B} serotonin receptors in addition to dopamine receptors, causing valvular heart disease

Frantz, S. *Nature* **2005**, *437*, 942-943.

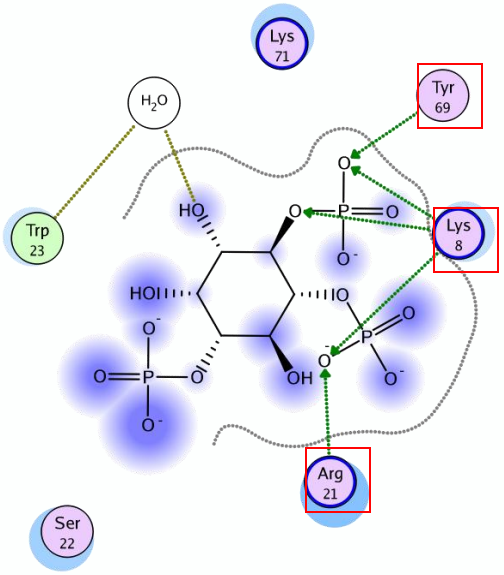
Keiser *et al*. *Nature* **2009**, *462*, 175-181.

Binding Site Representation



1b55

(IP binding)

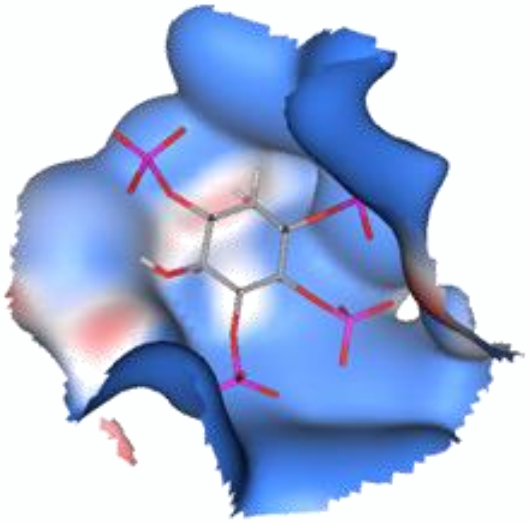


1btn

(IP binding)

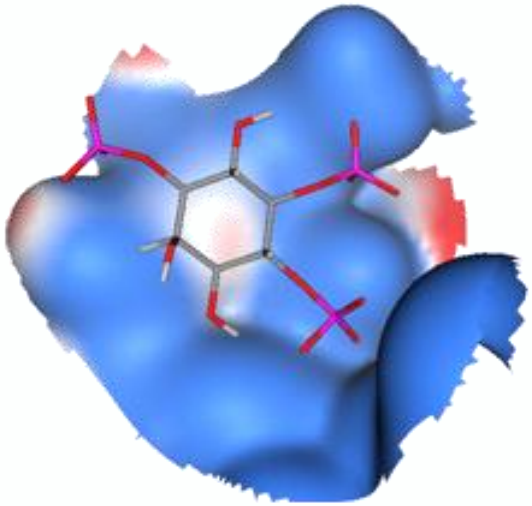
Low sequence conservation

Binding Site Representation



1b55

(IP binding)



1btn

(IP binding)

The EP mapped surfaces are similar

Binding Site Representation

– Molecular surfaces: sets of adjacent triangles mapped with property values at each vertex

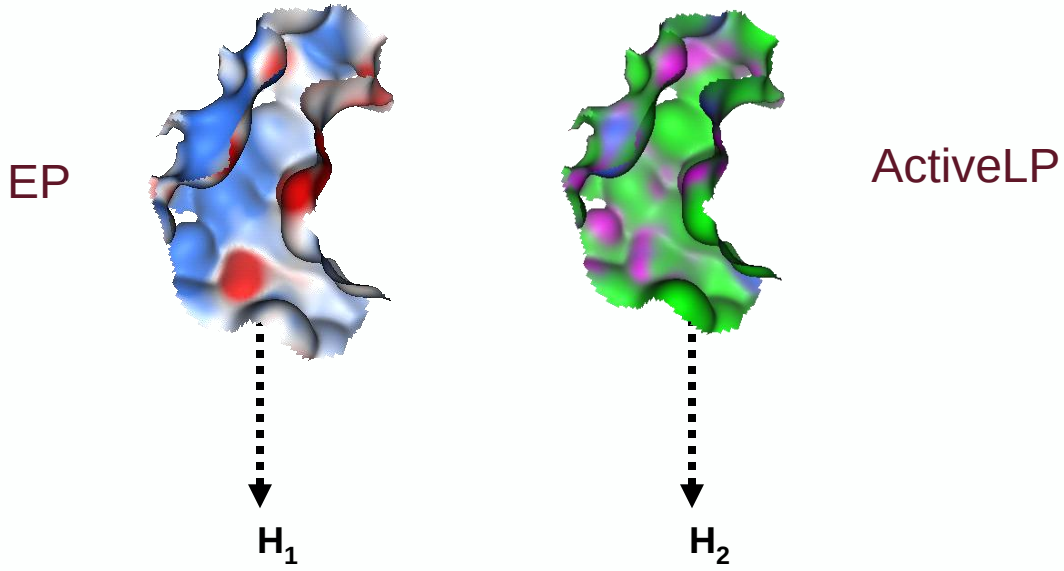
– A graphical binding site surface typically has 8000 to 12000 triangles

– Rigorous comparison involves matching each triangle: use of a clique detection algorithm that is computationally expensive (NP-hard)

– Slow for high-throughput similarity detection and global similarity search

Kinoshita, K.; Nakamura, H. *Protein Sci.* **2003**, *12*, 1589-1595.

Property-Encoded Shape Distributions

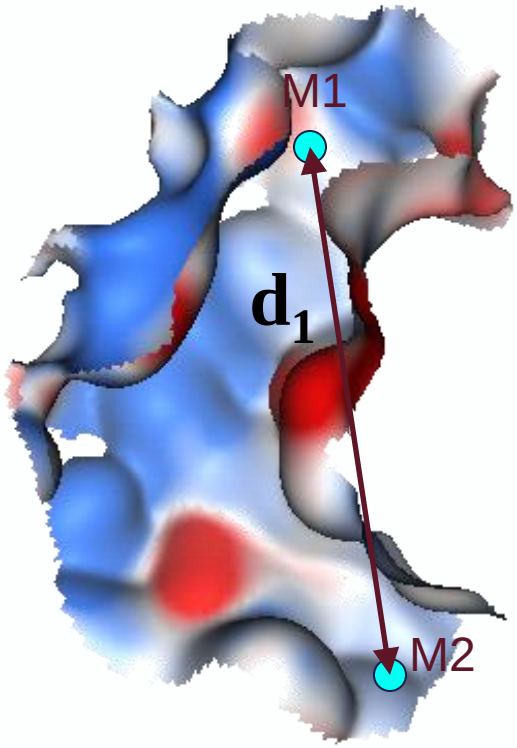


- Conversion of property distribution on surfaces to a string of numbers or signatures H_1, H_2, etc.

- Similarity between two binding sites is simply similarity between two signatures

Das, S.; Kokardekar, A.; Breneman, C. M. *J. Chem. Inf. Model.* **2009**, *49*, 2863-2872.

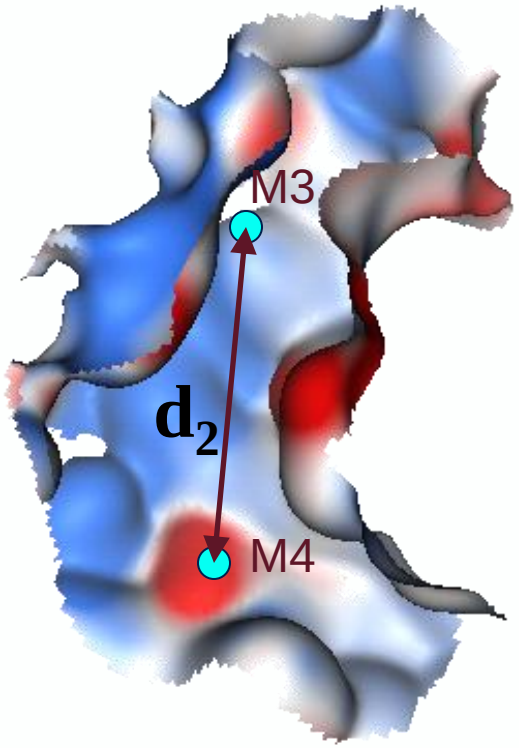
Property-Encoded Shape Distributions



- Large number of randomly selected pairs of points from the surface for convergence and binned by distance & property combinations

Osada R, Funkhouser T, Chazelle B, Dobkin D. Shape Distributions. *ACM Trans. Graph.* 2002, 21, 807-832

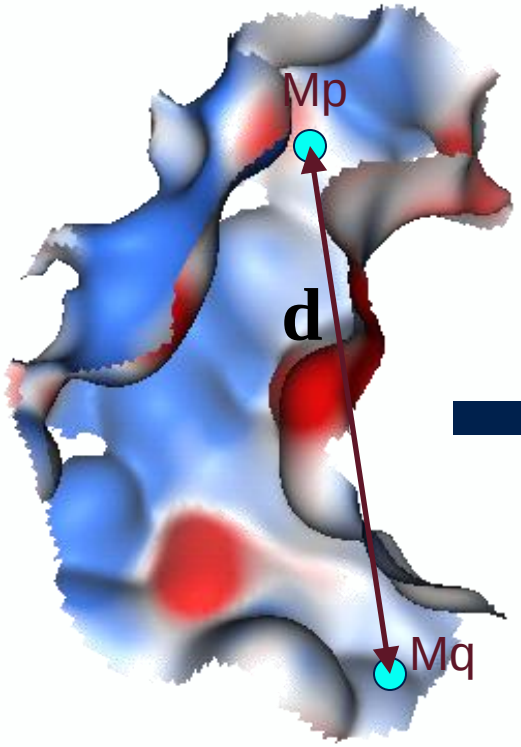
Property-Encoded Shape Distributions



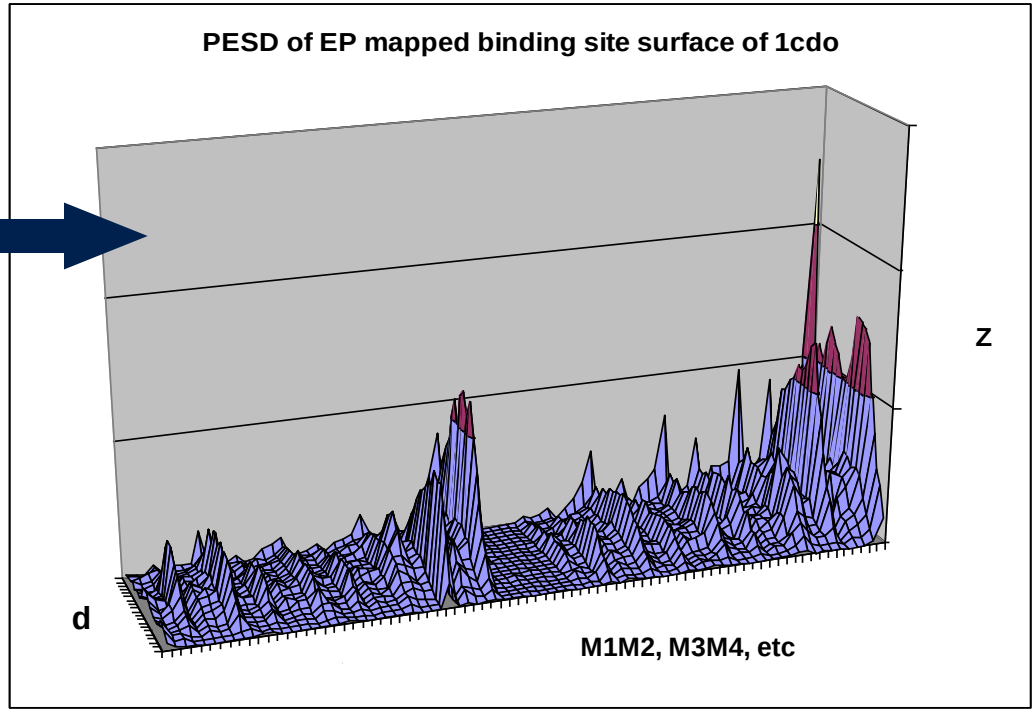
- Large number of randomly selected pairs of points from the surface for convergence and binned by distance & property combinations

Osada R, Funkhouser T, Chazelle B, Dobkin D. Shape Distributions. *ACM Trans. Graph.* 2002, 21, 807-832

Property-Encoded Shape Distributions



- Large number of randomly selected pairs of points from the surface for convergence and binned by distance & property combinations



Property-Encoded Shape Distributions

- Most property-encoded molecular surfaces are triangulated and color coded (representative of property and its magnitude) at each vertex
- ***To choose a surface point in an unbiased way:***
  1. Store triangles as array of cumulative areas
  2. Randomly choose a value ***x*** between 0 and total area
  3. For the triangle in the array having ***x*** within bounds of its cumulative area, calculate a point, such that:

$$P = (1-\sqrt{r_1})A + \sqrt{r_1}(1-r_2)B + \sqrt{r_1}r_2C$$

- The selected point is assigned the property magnitude of its nearest vertex

Property-Encoded Shape Distributions

- Signature comparison by chi-squared distance

$$d_{\chi^2}(H,K) = \sum_i \frac{(h_i - m_i)^2}{m_i} \quad ; \quad m_i = \frac{h_i + k_i}{2}$$

- Final distance score weighted sum of EP and ActiveLP distance

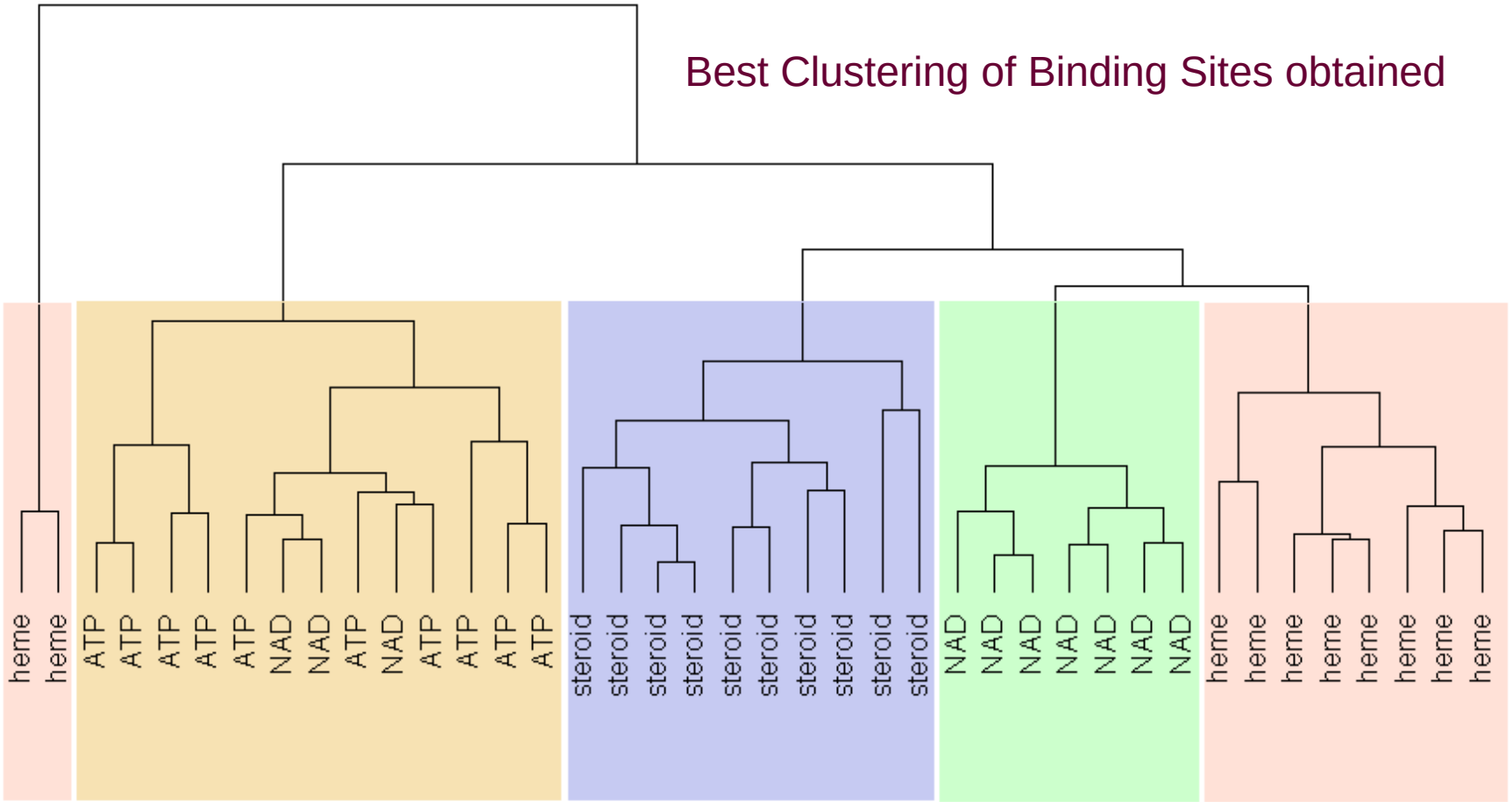
$$Score = d_{\chi^2_{EP}} + 0.7 d_{\chi^2_{ALP}}$$

Weight

- 10 to 15 seconds for query signature computation, >100,000 sites can be screened in ~5 minutes

Trial Clustering Analysis of Binding Site Signatures

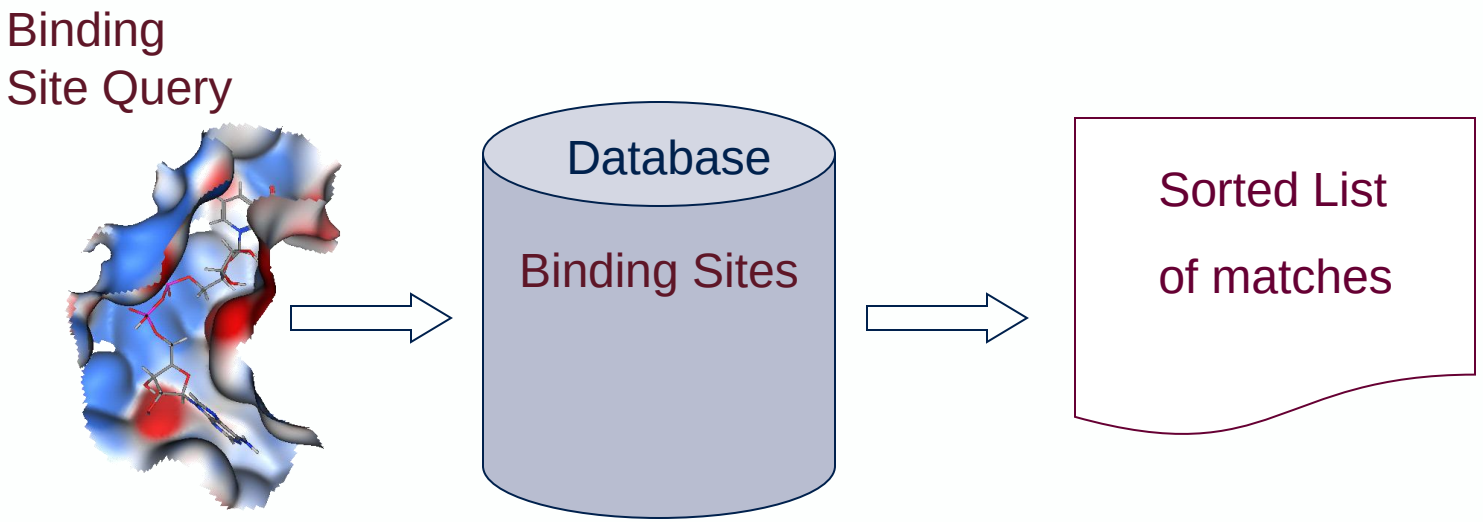
Best Clustering of Binding Sites obtained



Higher Accuracy than PocketMatch

(Yeturu *et al*. *BMC Bioinformatics*, **2008**, *9*, 543-559)

Virtual Screening with PESD



Screened on the PDBbind data set and the FINDSITE data set

Wang *et al. J. Med. Chem*. **2004**, *47*, 2977-2980.

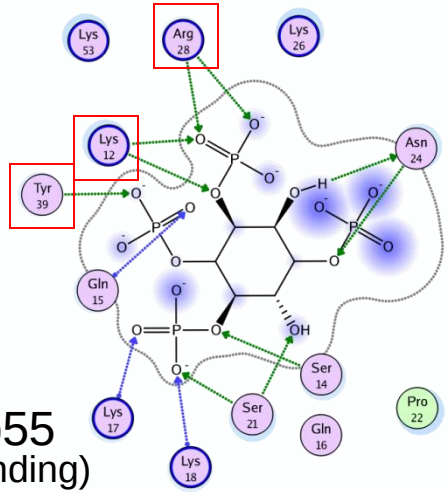
Brylinski, M.; Skolnick, J. *Proc. Natl. Acad. Sci. U.S.A*. **2008**, *105*, 129-134.

E.C. Similarity: Screening the PDBbind set

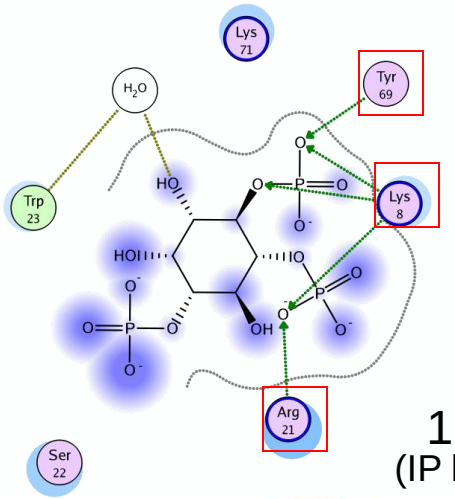
Ability of PESD to return a binding site with the same E.C. numbers in the top ranks of matched sites.

Top	1	3	1%	2%	5%
Positive (%)	79.5	85.1	87.9	89.7	92.5

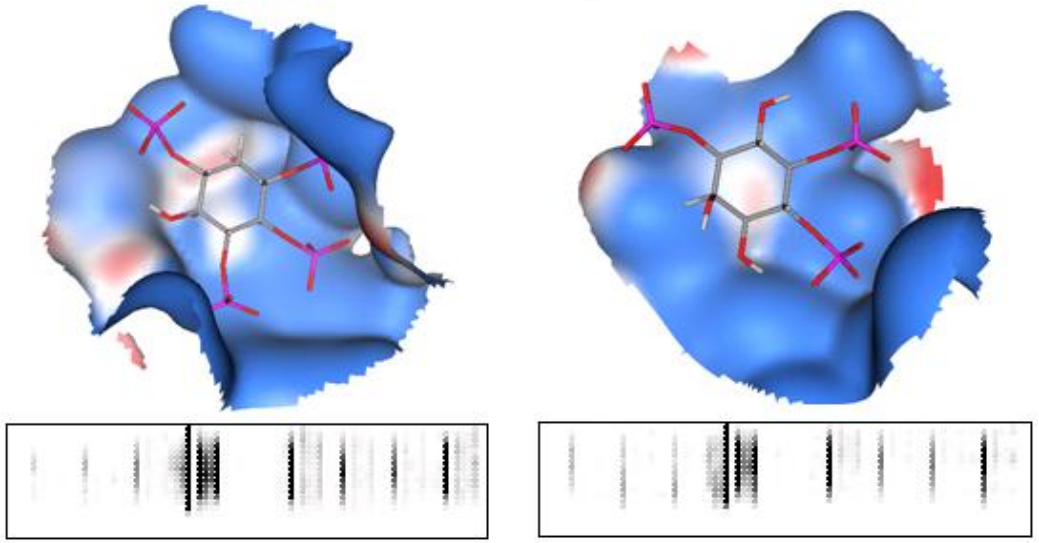
Case Studies



1b55
(IP binding)

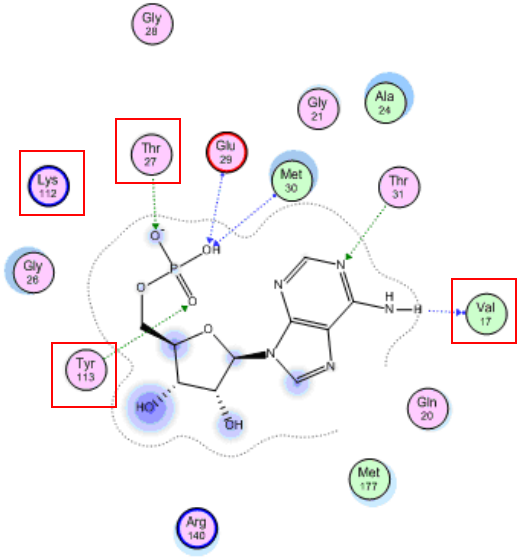


1btn
(IP binding)



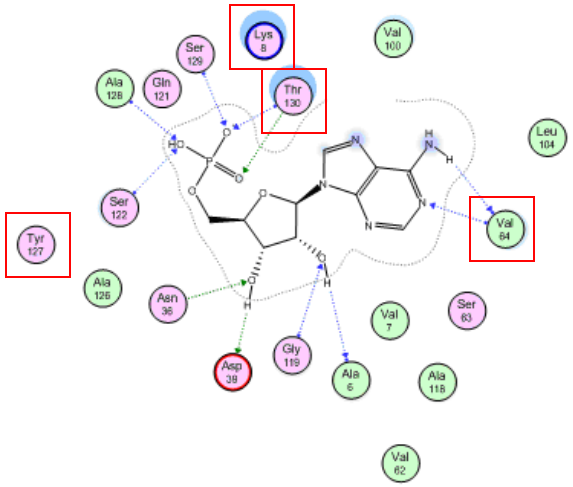
Low similarity of amino-acids

Case Studies



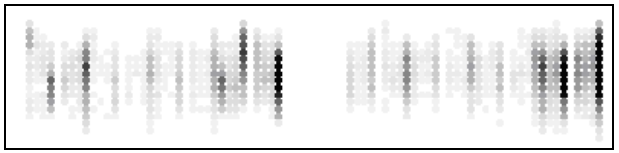
1fbp

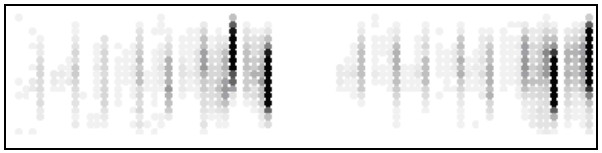
Fructose-1,6-bisphosphatase



1o97

Flavoprotein

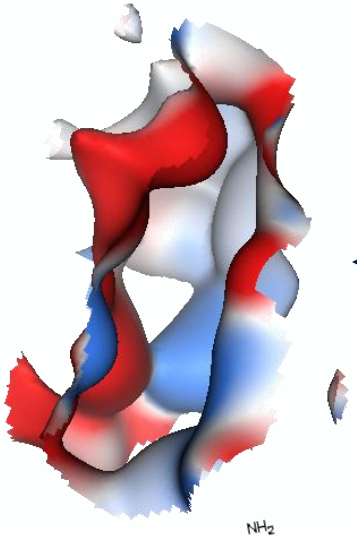


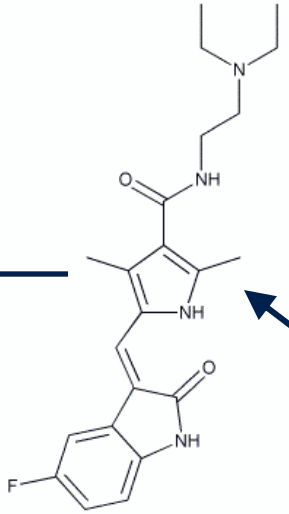


Low similarity of amino-acids

Case Studies

1cdk (cAMP dependent Kinase)

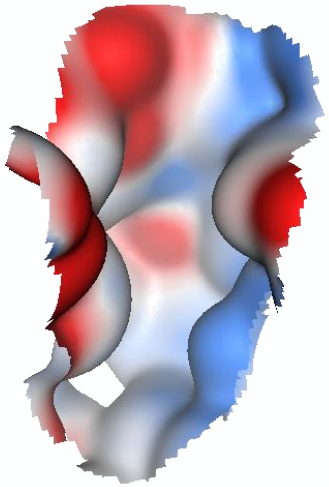


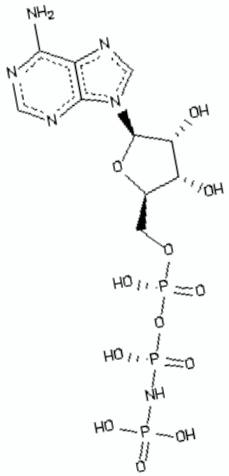


SU11248

(Sunitinib, Sutent)

2jav (Nek2 Kinase)





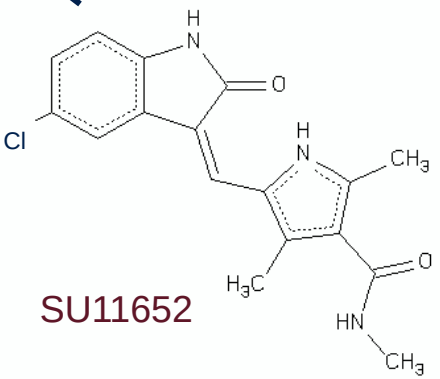
ANP

ROCS ligand similarity

0.488 : Shape

0.660 : Combo

Ligands dissimilar

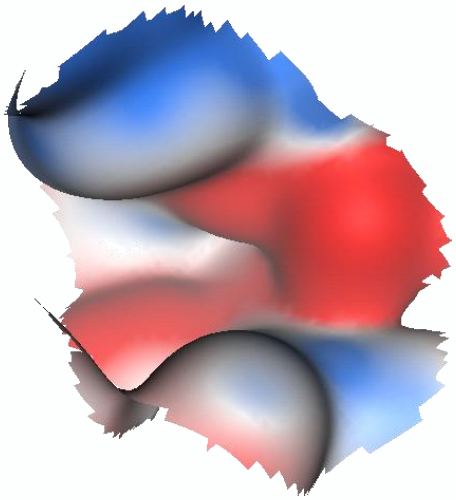


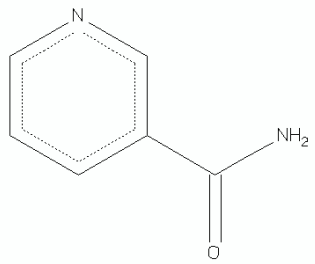
SU11652

Case Studies

1isi

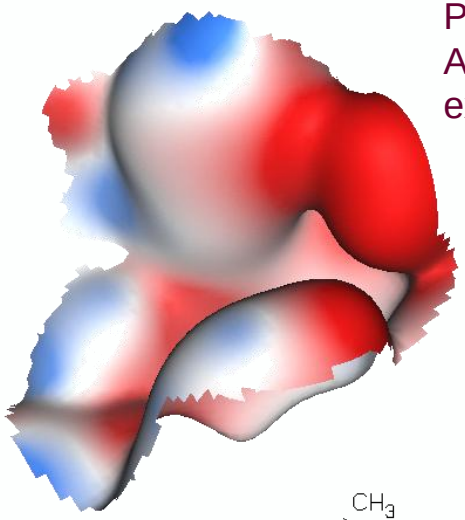
BST-1

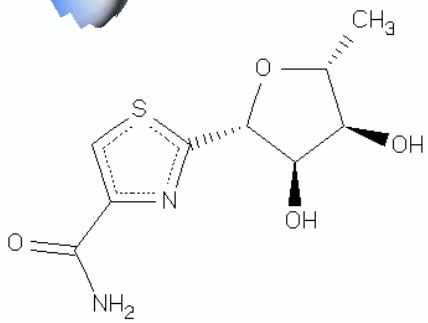




1aer

Pseudomonas Aeruginosa exotoxin



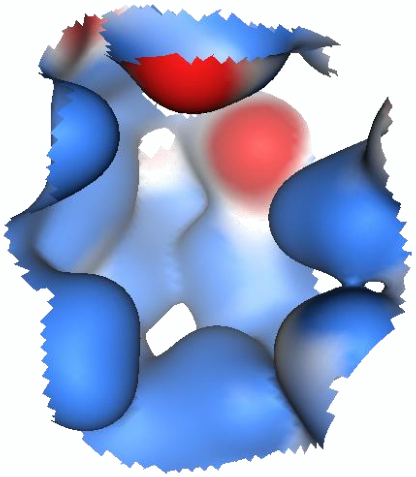


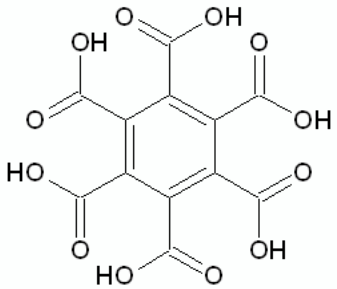
Ligand sub-structural similarity from binding site similarity

Case Studies

1bq4

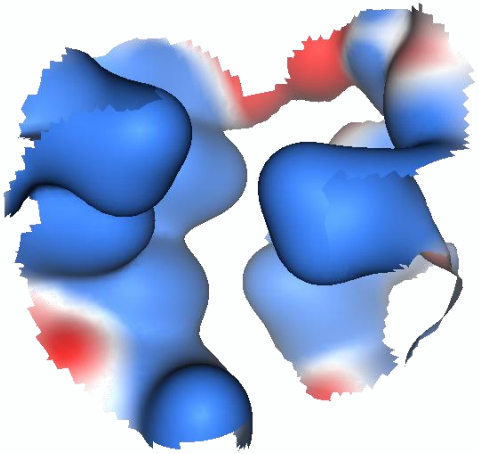
Phosphoglycerate mutase

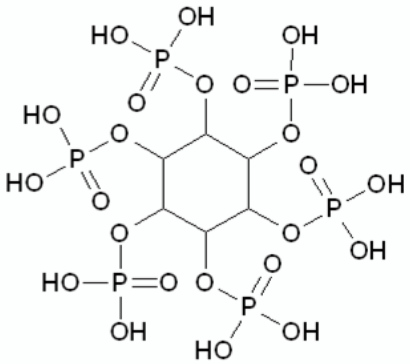




2fvv

Human diphosphoinositol polyphosphate phosphohydrolase 1



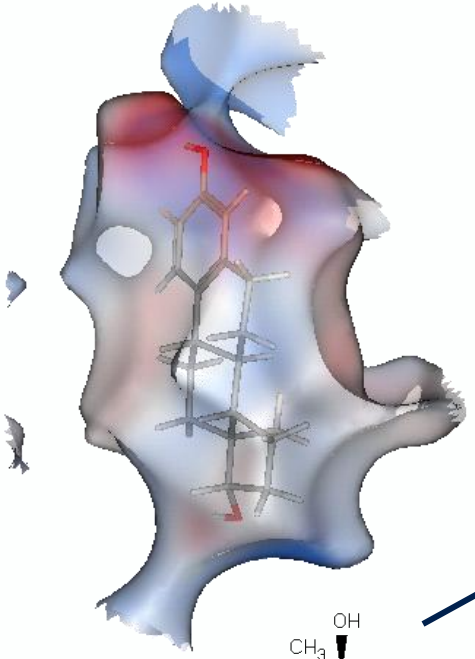


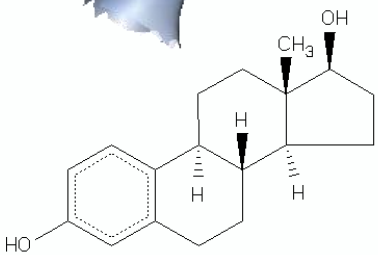
Ligand sub-structural similarity from binding site similarity

Case Studies

1lhu (Sex hormone binding globulin)

All beta proteins



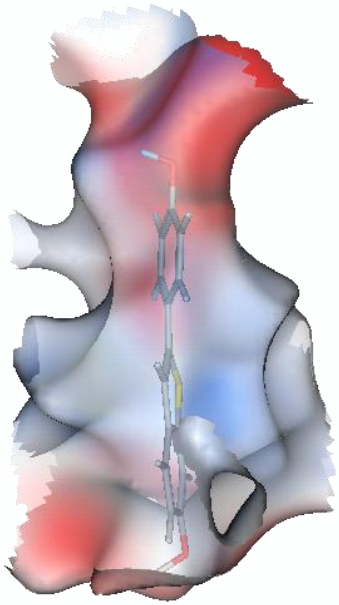


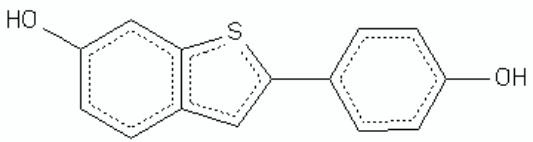
Bound Estradiol

Estradiol binding

1gwq (Estrogen receptor)

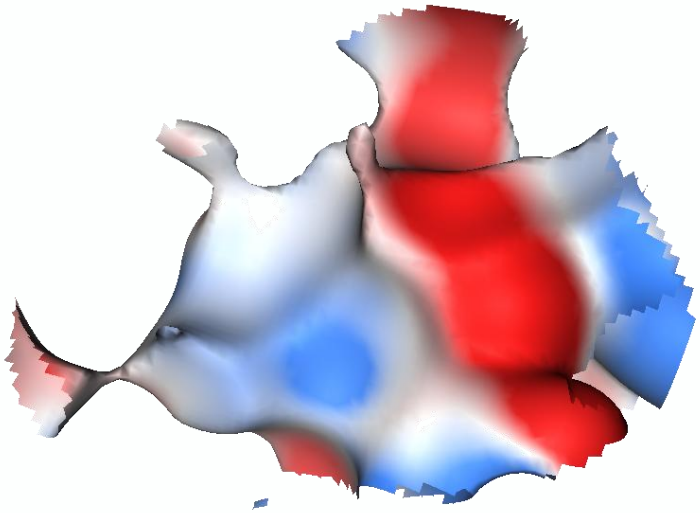
All alpha proteins

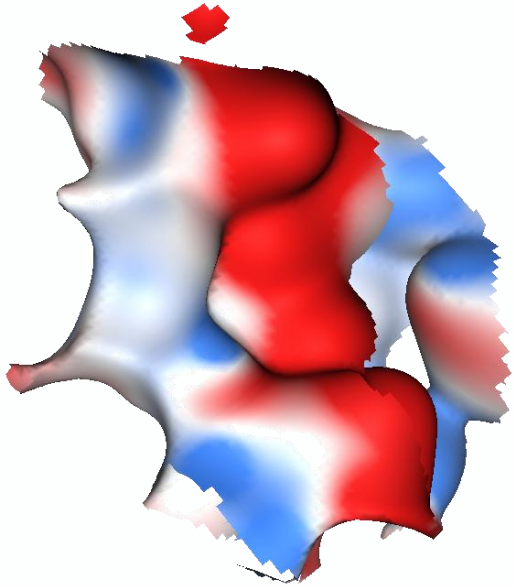




Bound Raloxifene Core

Case Studies

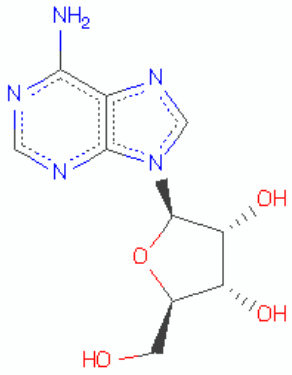




1fmo (cAMP dep protein kinase)

Staurosporine binding

1nvs (checkpoint kinase Chk1)



MACCS structural keys

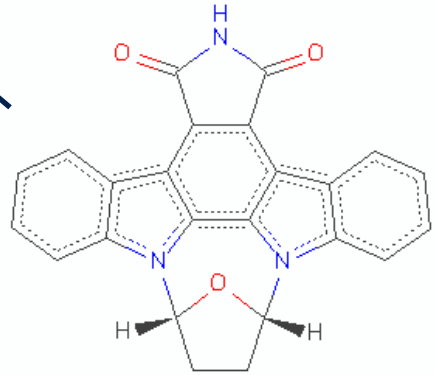
TC = 0.5

ROCS

Shape:0.5 Combo:0.7

Rank in ROCS **Tversky$_q$**: Top 3%

Rank in PESD: Top 1%



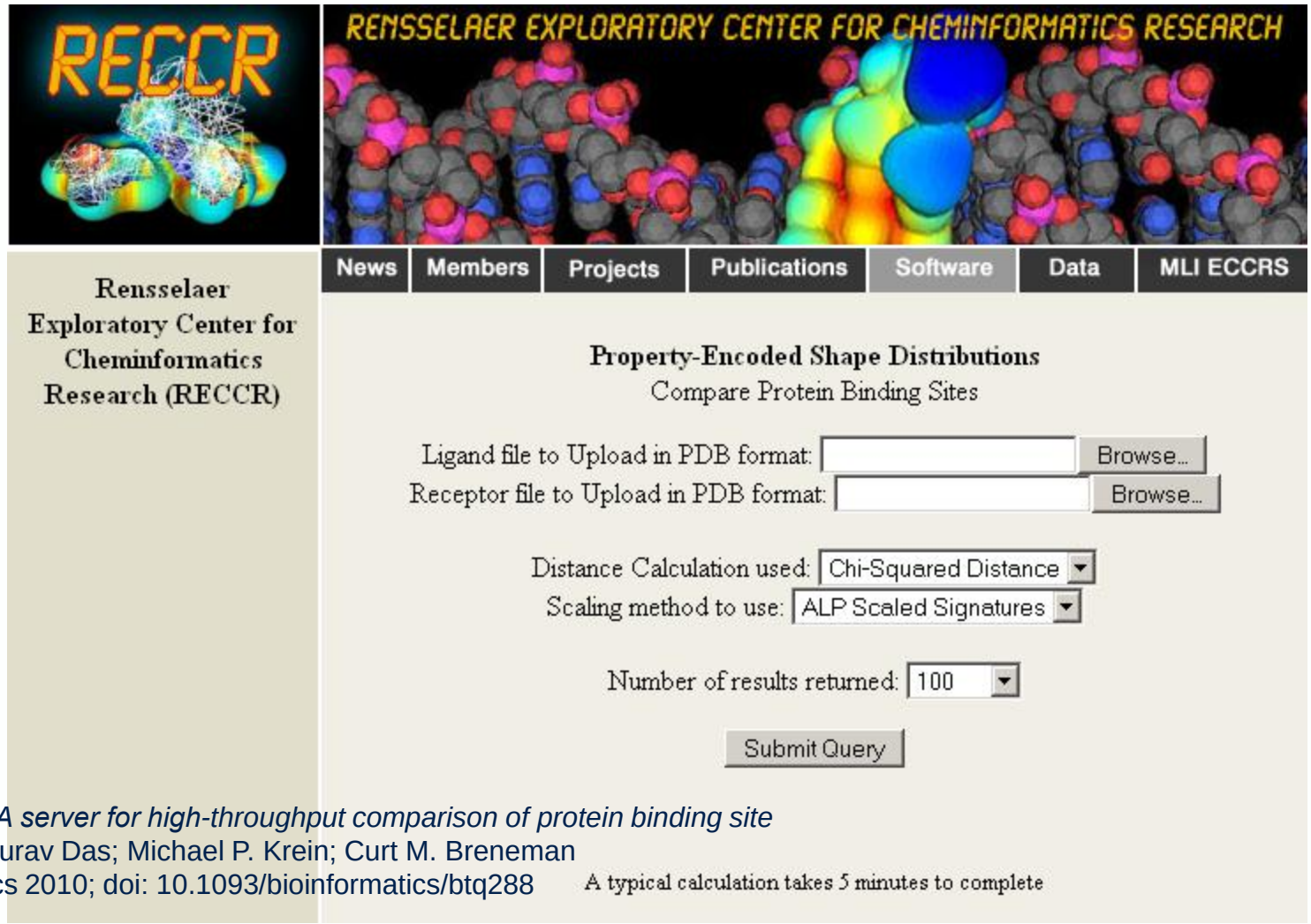
Case Studies

Similarity among ATP binding sites : cross-reactivity of a promiscuous binder – ATP itself

Rank	Query: 1cdk		
	PDB	Ligand	
1	1mzv	AMP	
2	1b62	ADP	
3	1xdn	ATP	RNA-editing Ligase
4	2c5s	AMP	
5	1q5h	DUD	
6	1ia9	ANP	TranRecPot-related P
7	1jm6	ADP	Pyruv. De. Kinase
8	1aux	SAP	
9	1zxm	ANP	Topoisomerase II
10	1byk	T6P	

cAMP-dep Pr. Kinase

PESD-serv http://reccr.chem.rpi.edu/Software/pesdserv/



"PESDserv: A server for high-throughput comparison of protein binding site surfaces" Sourav Das; Michael P. Krein; Curt M. Breneman
Bioinformatics 2010; doi: 10.1093/bioinformatics/btq288

PESD-serv http://reccr.chem.rpi.edu/Software/pesdserv/

Ligand ID

PESD Score

Ligand Chain ID

PDB

PDB entry title

PDB	Ligand ID	PESD Score	Ligand Chain ID	PDB entry title
1a52	EST	1601.47	A 1	ESTROGEN RECEPTOR ALPHA LIGAND-BINDING DOMAIN COMPLEXED TO ESTRADIOL
1gwr	EST	3629.29	A 600	HUMAN OESTROGEN RECEPTOR ALPHA LIGAND-BINDING DOMAIN IN COMPLEX WITH 17BETA-OESTRADIOL AND TIF2 NRBOX3 PEPTIDE
2g44	T3O	3787.73	B 701	HUMAN ESTROGEN RECEPTOR ALPHA LIGAND-BINDING DOMAIN IN COMPLEX WITH OBCP-1M-G AND A GLUCOCORTICOID RECEPTOR INTERACTING PROTEIN 1 NR BOX II PEPTIDE
1ere	EST	3791.36	A 600	HUMAN ESTROGEN RECEPTOR LIGAND-BINDING DOMAIN IN COMPLEX WITH 17BETA-ESTRADIOL REVDAT 2 2 3 TITLE HETATM MTRIX JRNL
1ere	EST	3832.95	B 600	HUMAN ESTROGEN RECEPTOR LIGAND-BINDING DOMAIN IN COMPLEX WITH 17BETA-ESTRADIOL REVDAT 2 2 3 TITLE HETATM MTRIX JRNL
1zky	689	3973.70	A 700	HUMAN ESTROGEN RECEPTOR ALPHA LIGAND-BINDING DOMAIN IN COMPLEX WITH OBCP-3M AND A

PESD Limitations

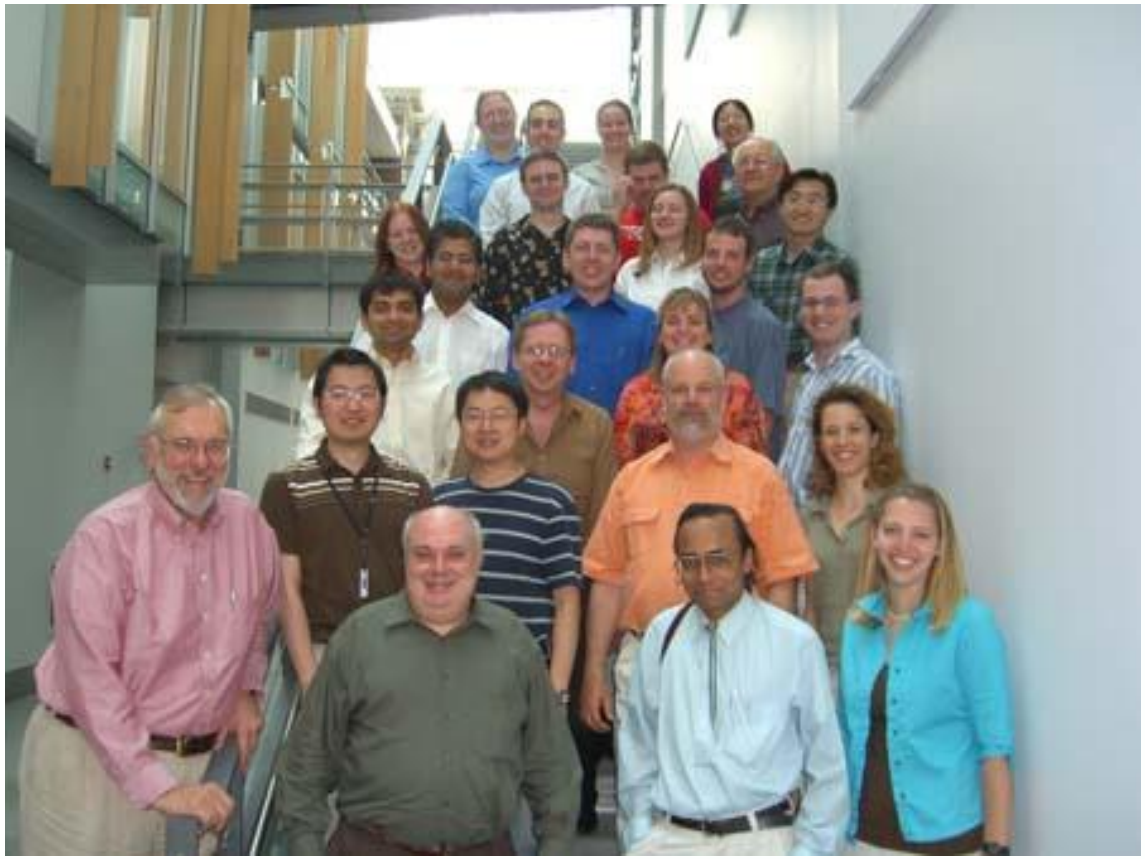
- Global matching results in lower similarity when relevant binding sites differ greatly in size
- Not suitable when flexible ligands bind to sites of significantly different shapes
- Requires development for making comparisons with unbound binding sites

Currently a postdoc…

Summary

- **RS_Predictor**
  - Advantages of Multiple-Instance Ranking
  - Use of Hierarchical Electronic and Topological Descriptors
  - Implicit encoding of CYP geometry for nine isozymes
  - RS_Predictor + Explicit Docking (Taowei Huang)
  - RS_Predictor + SmartCYP + Explicit Docking (soon)
- **PESD**
  - Encoding of binding site shape and features
  - Independent of sequence
  - Allows off-target interactions to be identified

The RECCR Community



http://reccr.chem.rpi.edu

ACKNOWLEDGMENTS

- Current and Former members of the RECCR/DDASSL group
  - Breneman Research Group (RPI Chemistry)
    - N. Sukumar
    - M. Sundling
    - Min Li
    - Long Han
    - Jed Zaretski
    - Taowei Huang
    - Theresa Hepburn
    - Mike Krein
    - Steve Mulick
    - Shiina Akasaka
    - Hongmei Zhang
    - C. Whitehead (Pfizer Global Research)
    - L. Shen (BNPI)
    - L. Lockwood (Syracuse Research Corporation)
    - M. Song (Synta Pharmaceuticals)
    - D. Zhuang (Simulations Plus)
    - W. Katt (Yale University chemistry graduate program)
    - Q. Luo (J & J)
  - Embrechts Research Group (RPI DSES)
  - Tropsha Research Group (UNC Chapel Hill)
  - Bennett Research Group (RPI Mathematics)
- Collaborators:
  - Brinson Advanced Materials Lab – (Northwestern)
  - Schadler Research Group (RPI Materials Engineering)
- Funding
  - Office of Naval Research (ONR)
  - Lockheed-Martin (LMCO)
  - NIH (GM047372-07)
  - NIH (1P20HG003899-01)
  - NSF (BES-0214183, BES-0079436, IIS-9979860)
  - GE Corporate R&D Center
  - Ford-Boeing Alliance
  - NSF NIRT Program
  - Chemical Computing Group (CCG)