Cheminformatics Approaches for Hazard Identification & Characterization

William (Bill) Welsh

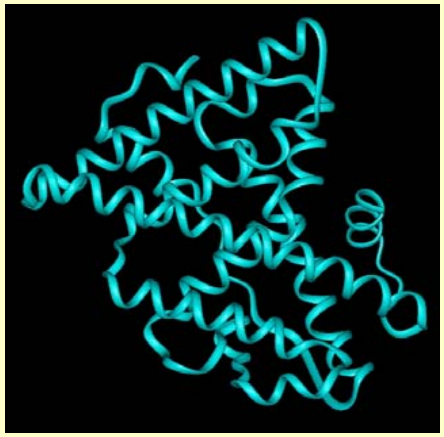
welshwj@umdnj.edu

Environmental Bioinformatics & Computational Toxicology Center

www.ebCTC.org

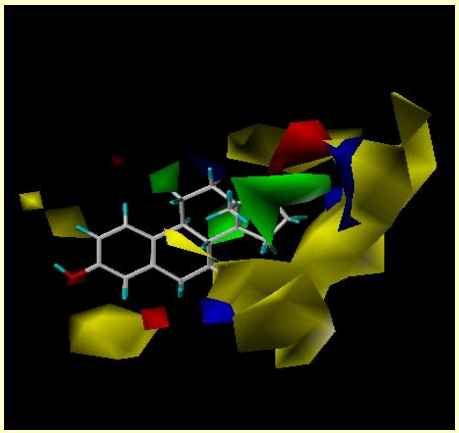
Integrated Approach

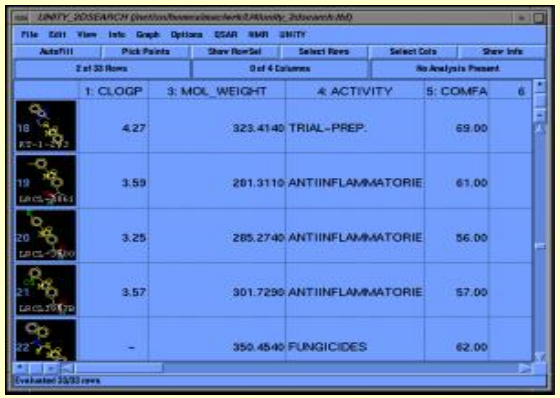
Receptor-based Approaches

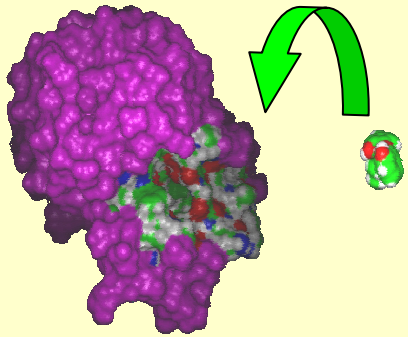


Predictive Molecular Toxicology

Ligand-based Approaches



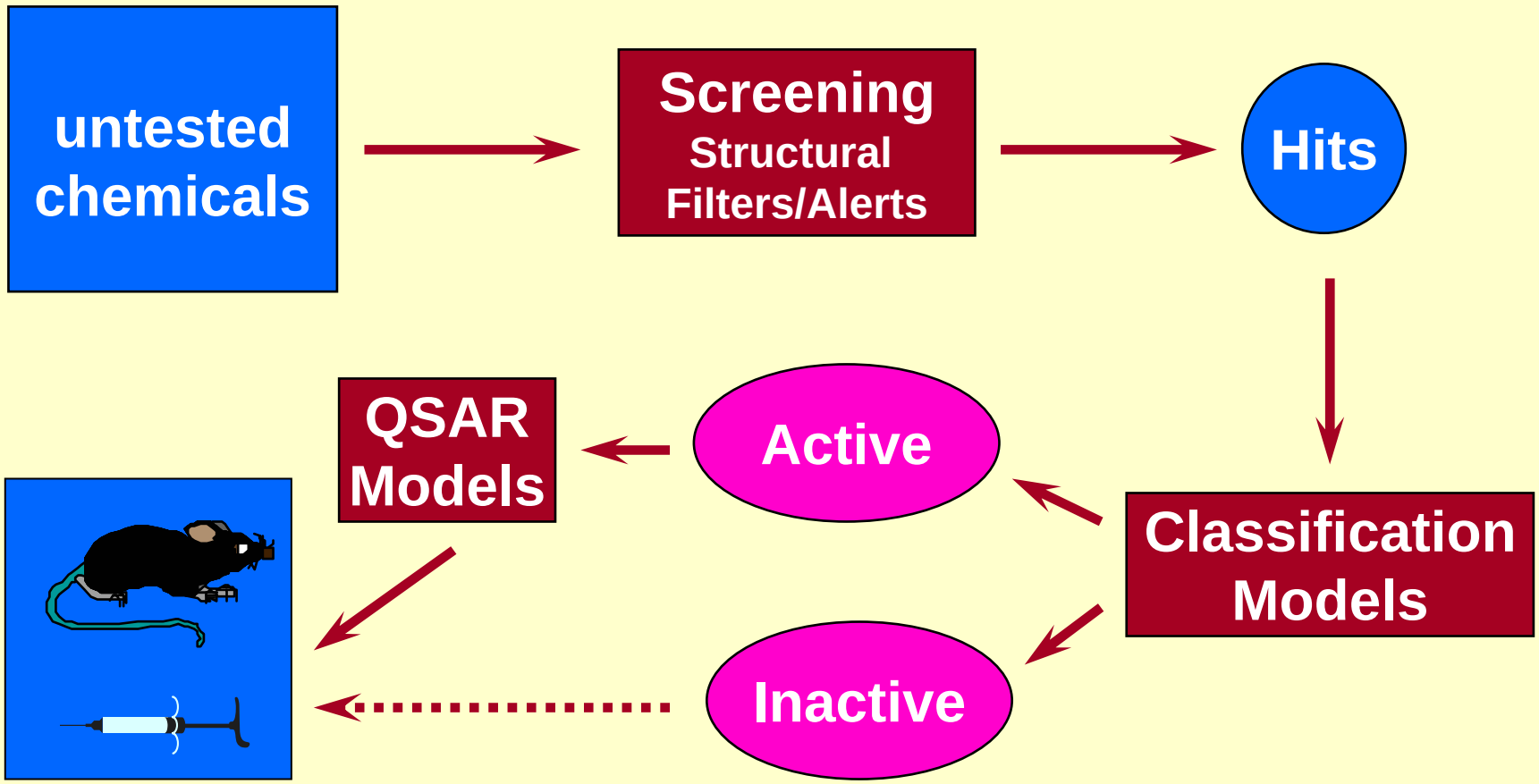




Virtual Screening

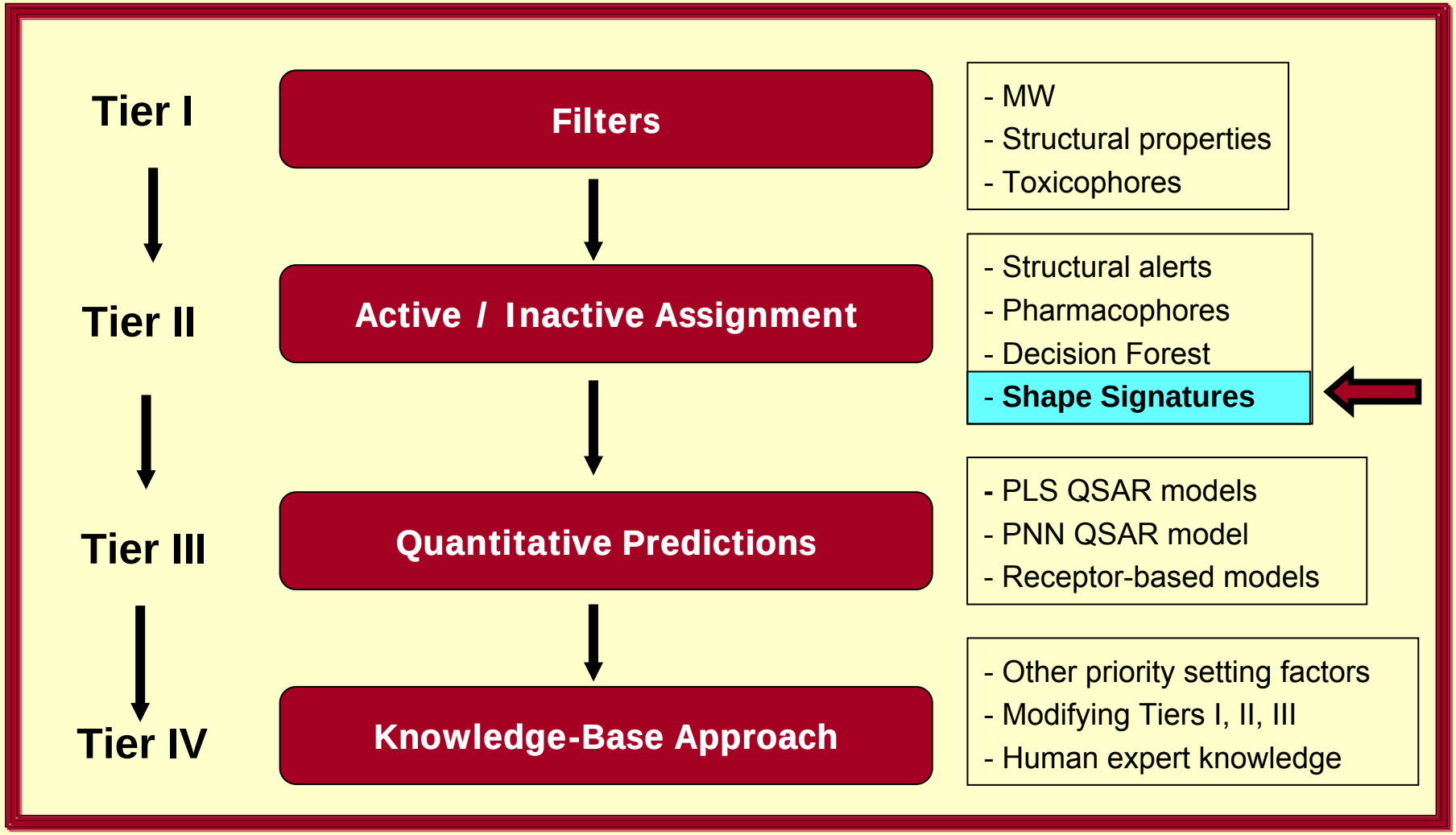
Computational Screening Paradigm

- Priority Setting -



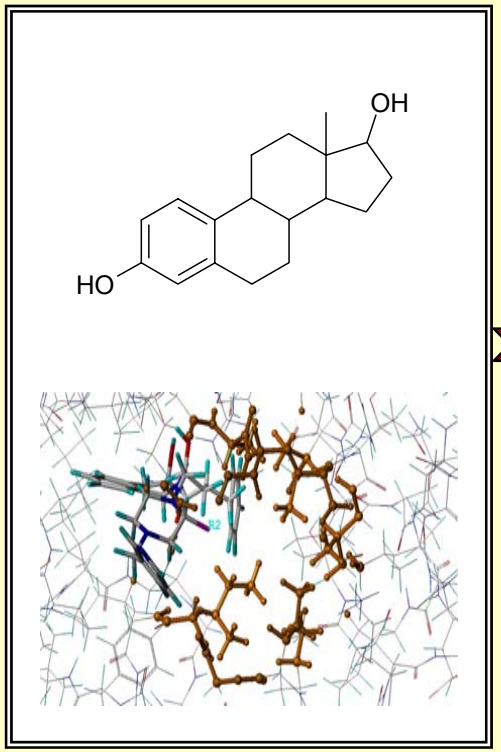
Schematic of Hierarchical Screening Framework

- addresses the need to minimize *false negatives* and *uncertainties* -



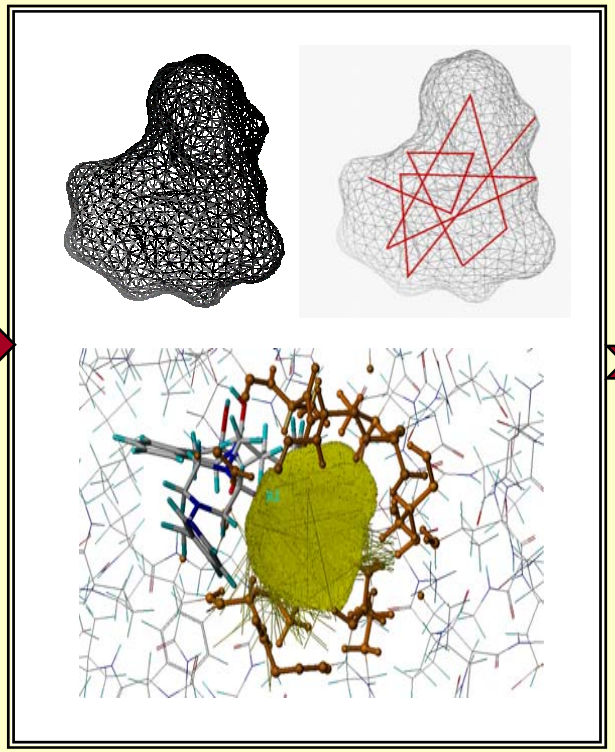
Shape Signatures Tool

START



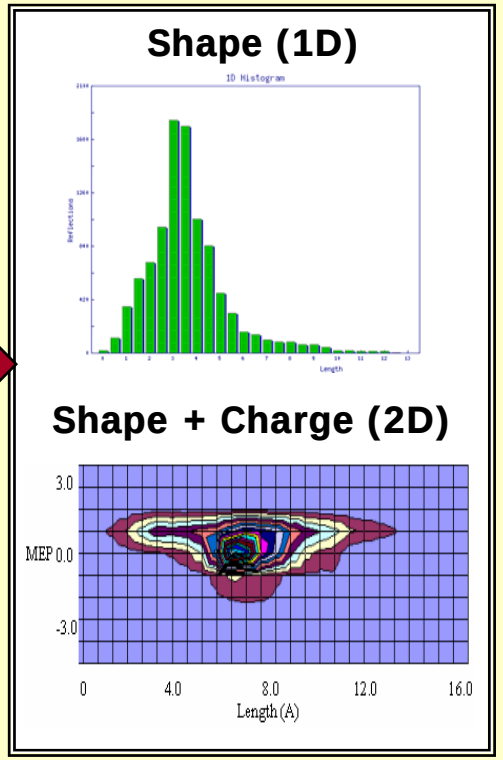
Small molecule or Protein binding pocket

PROCESSING



Ray tracing to generate the raw data

OUTPUT

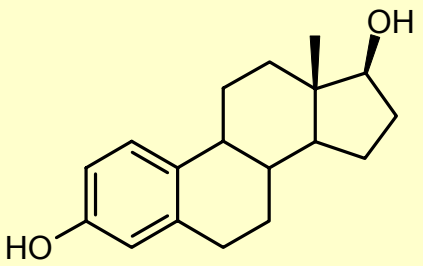


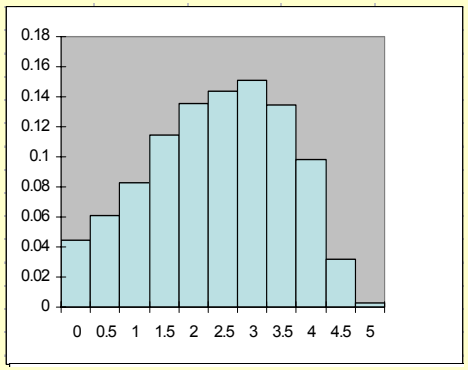
1D and 2D Shape Signatures

Shape Signatures Tool

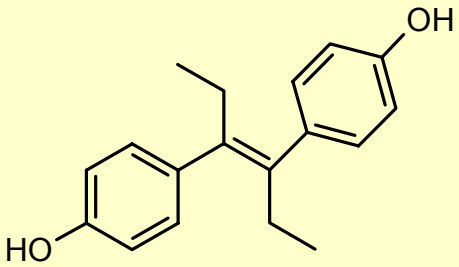
molecules are compared by subtracting their histograms

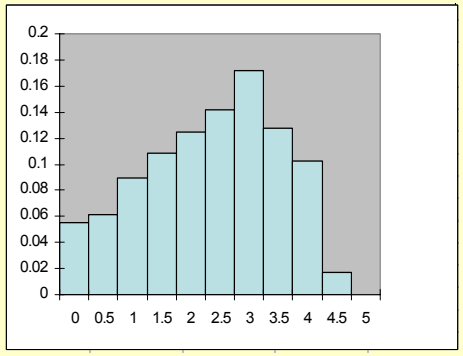
17β-estradiol





DES





***Diff* = 0.082**

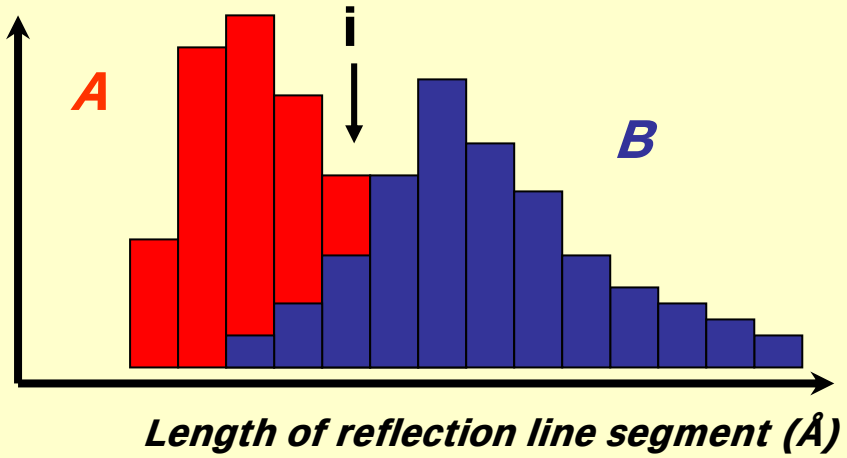
Small *Diff* value means that two molecules have similar shape and polarity

Shape Signatures Tool: Scoring Schemes

compute the difference between two normalized histograms representing molecules *A* and *B*

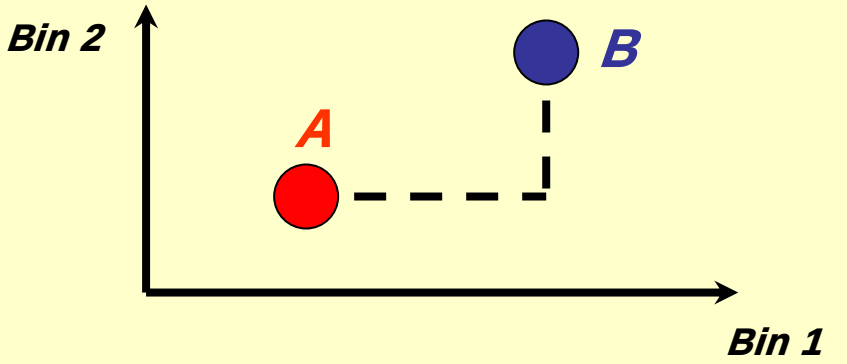
1. χ^2 score

$$\Delta_{AB} = \sum_{i}^{\text{over bins}} (A_i - B_i)^2 / (A_i + B_i)$$



2. Manhattan metric score

$$\Delta_{AB} = \sum_{i}^{\text{over bins}} | A_i - B_i |$$

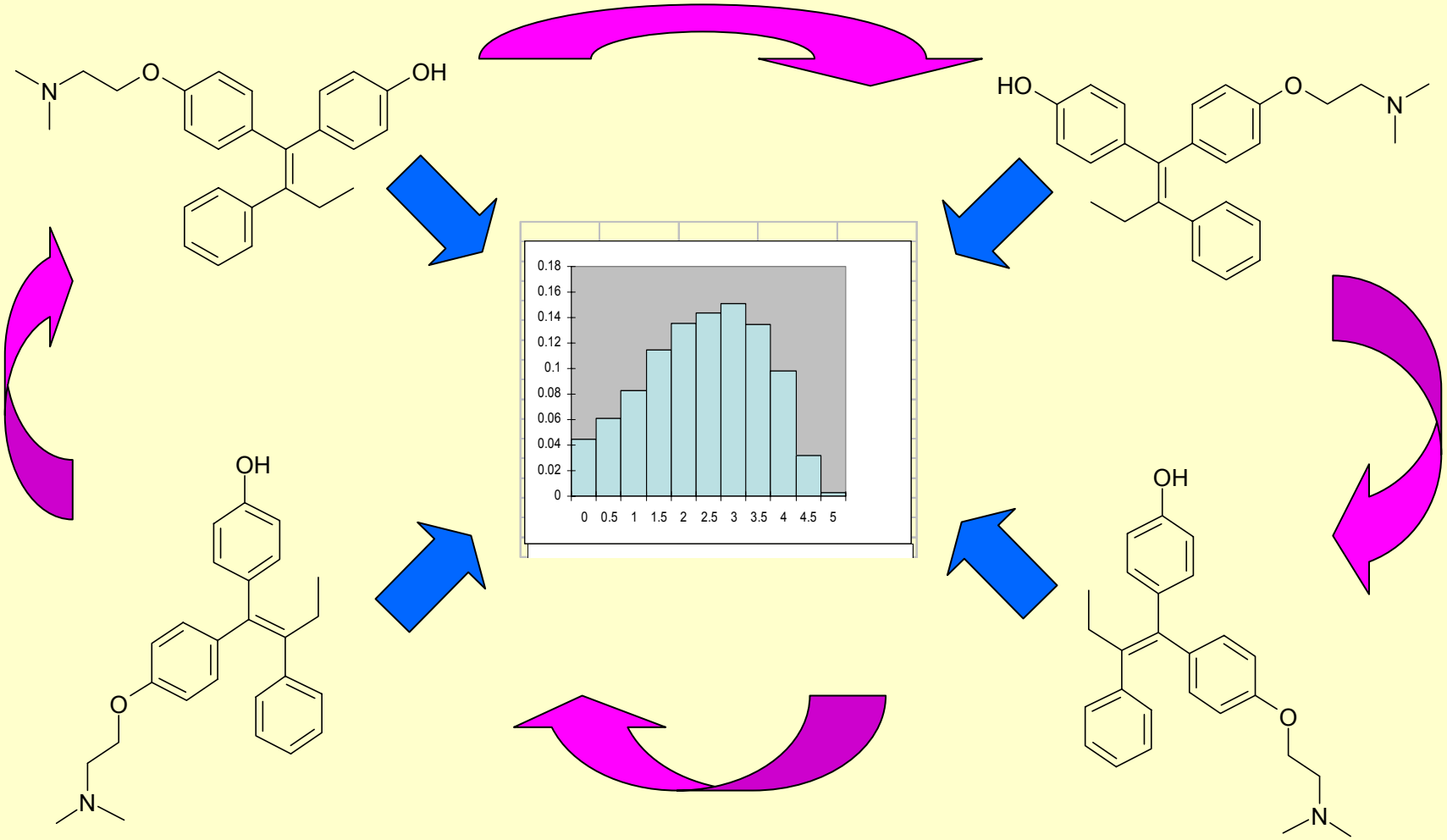


$\Delta_{AB} = 0$ ⟹ ***A*** **and** ***B*** **are identical molecules (complete overlap)**

$\Delta_{AB} = 2$ ⟹ ***A*** **and** ***B*** **are entirely different (no overlap)**

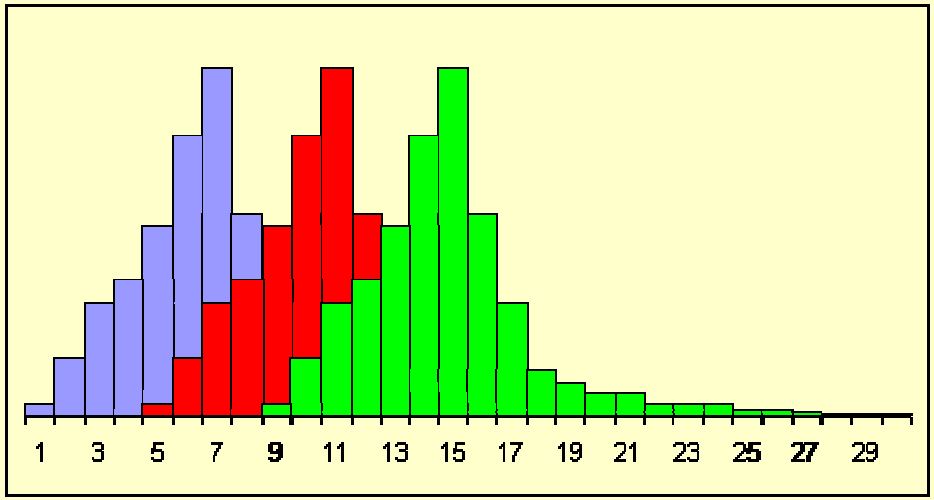
Shape Signatures are Rotationally Invariant

- insensitive to position/orientation -

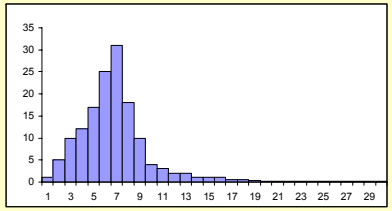


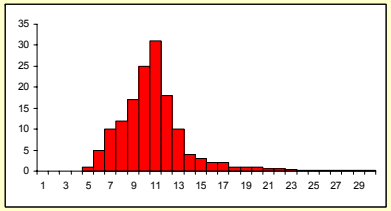
Shape Signatures:

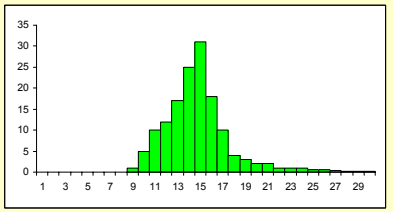
- same shape, but different sizes (volumes) -

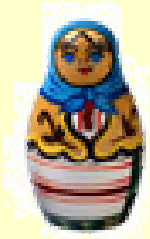


shift along x-axis

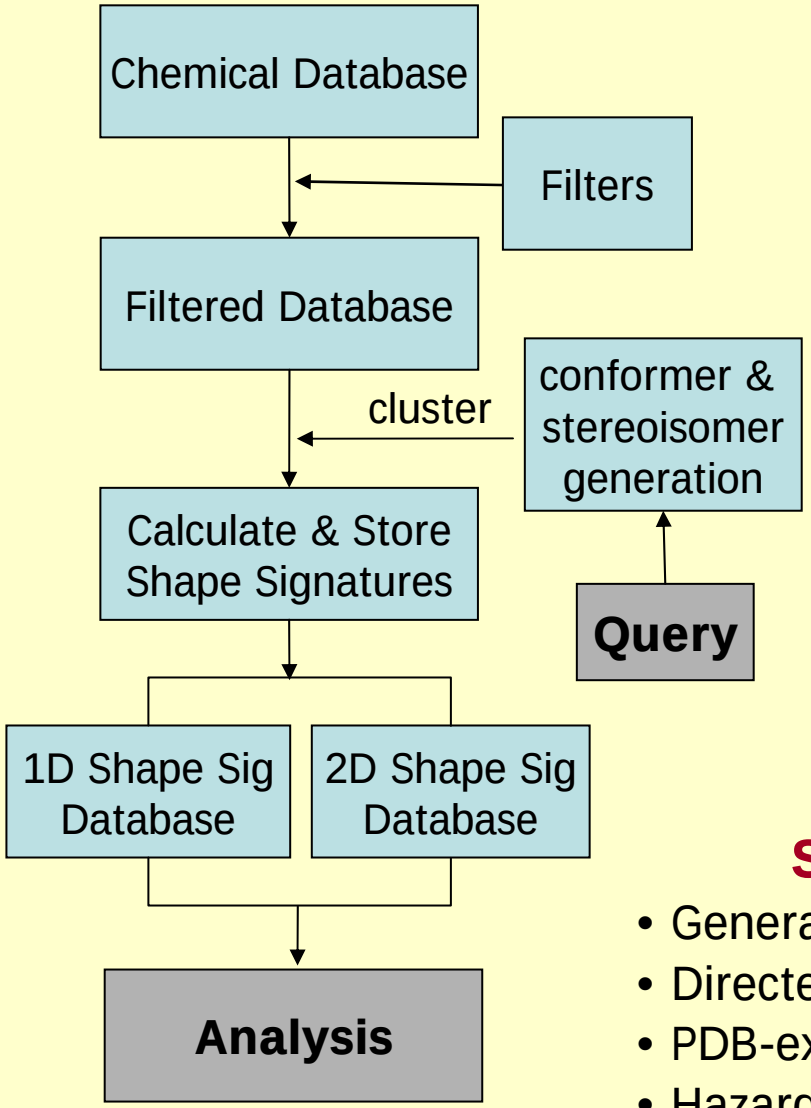




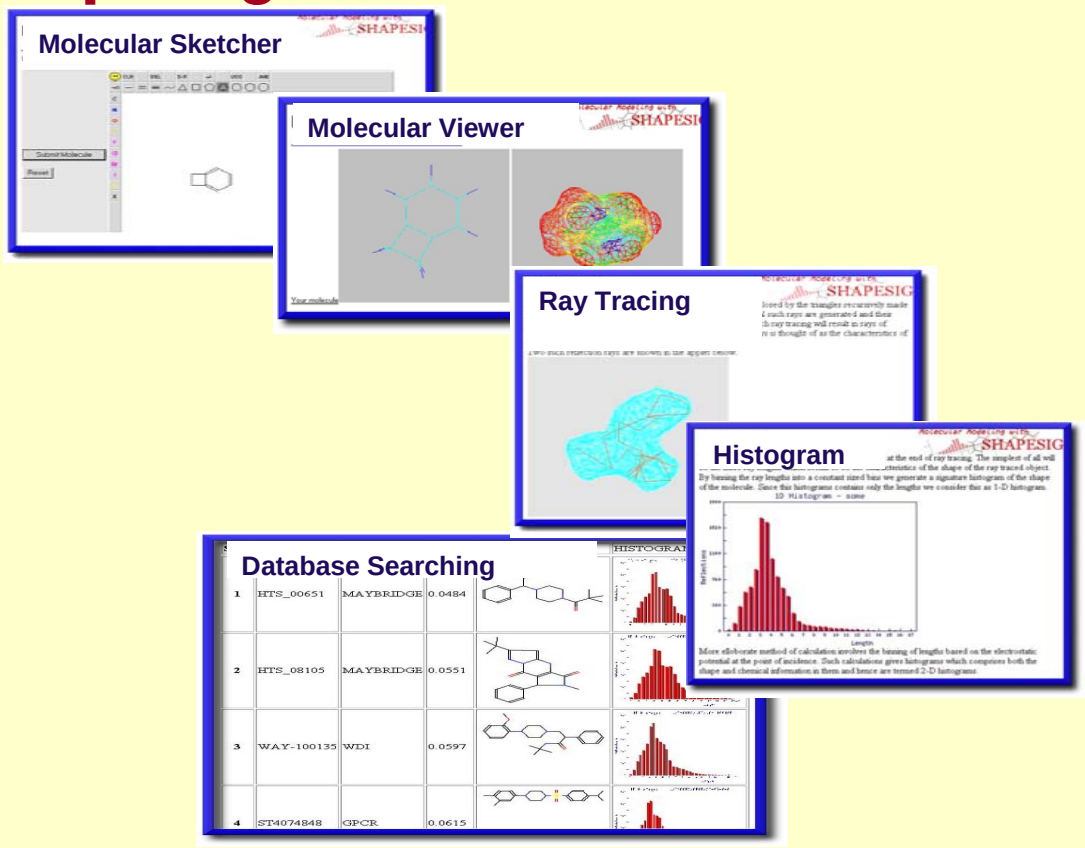




Flowchart



Shape Signatures User Interface



Shape Signature Databases

- General Database >3 million vendor chemicals
- Directed libraries of kinase, GPCR, and NR ligands
- PDB-extracted ligand database
- Hazardous Chemicals (EDCs, H_2O CCLs, DSSTox, CWAs)

Shape Signatures Database

Home
- Method
- Applicability
- Ongoing Projects
- Feedback
- References

Small Molecules Shape Signatures

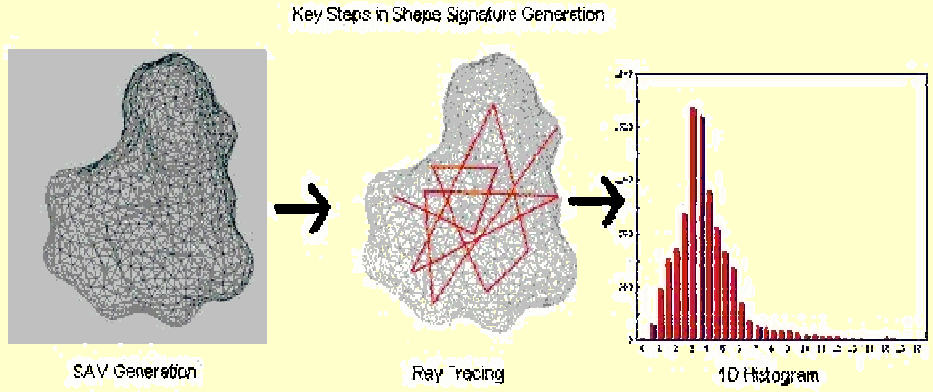
PDB Ligands Shape Signatures

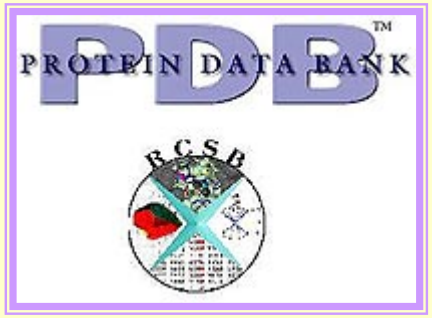
Contact Us

Last Updated: 09-07-2006

Description

E-lab, or *Shape Signatures*, a recently developed computational tool in Dr. William J. Welsh's lab, employs a customized ray-tracing algorithm to explore the volume enclosed within the surface of a molecule. It then processes the output to construct compact representations (i.e., signatures) of molecular shape, polarity, and other bio-relevant properties. Its underlying principle is that the molecular shape and surface charge of a chemical are fundamental determinants of its biological activity. To facilitate use by the general scientific community, the current version of Shape Signatures tool features numerous pre- and post-processing capabilities. For example, the computational engine has been embedded in an intuitive graphics user interface (GUI) that can be accessed through the internet. Molecules can be sketched or uploaded, and results from shape comparisons can be viewed in real time or stored. The program accepts molecular inputs in a wide variety of file formats, and procedures are implemented to handle conformational flexibility and stereochemistry.





Extract Ligands

High quality crystal structures
(resolution < 2.5 Å)

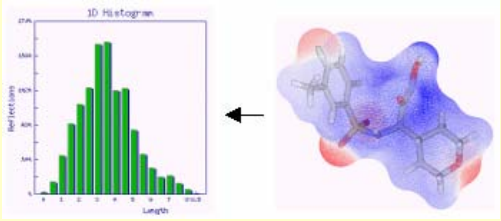
Preliminary Ligand Database

Remove metal ions, salts, redundant structures

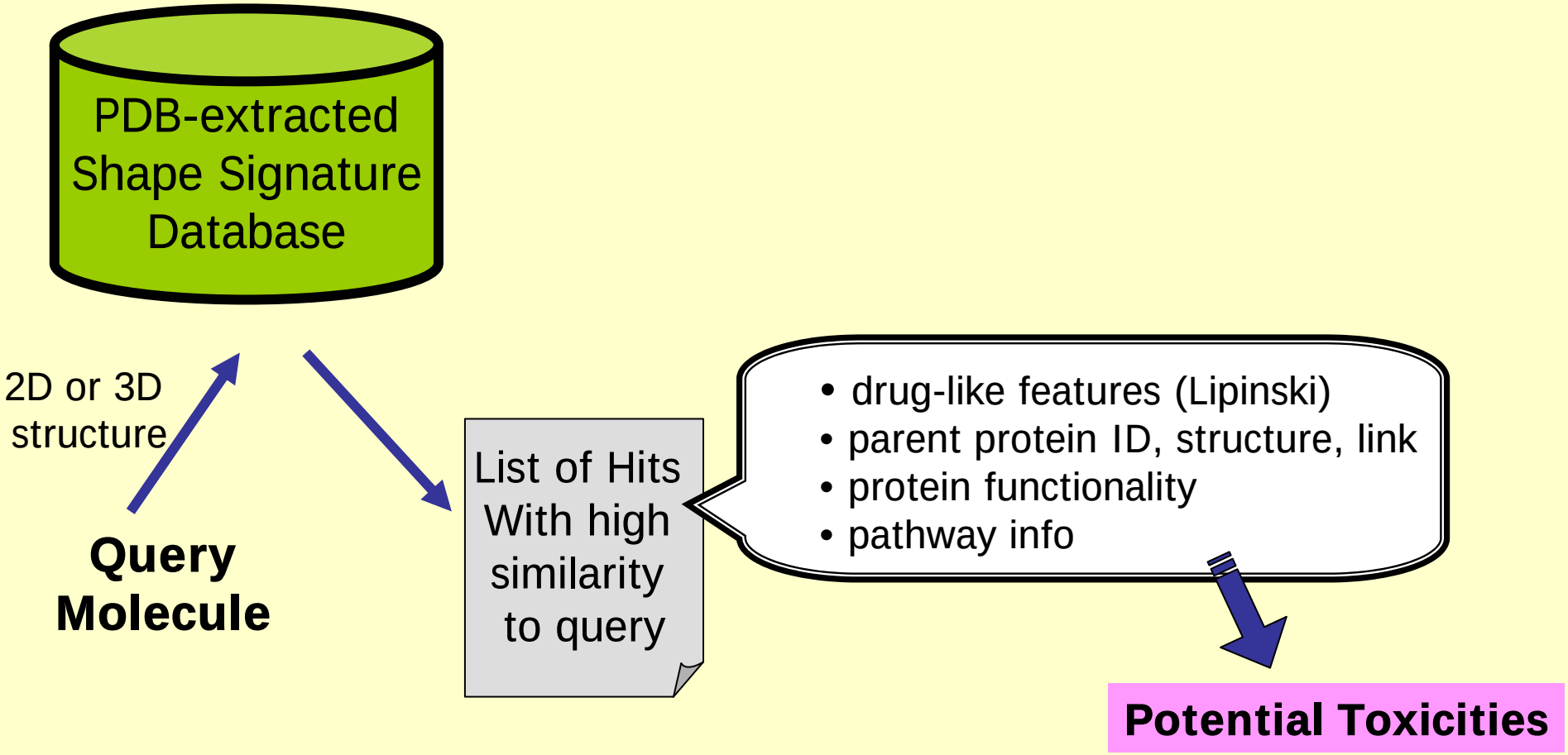
Filter

Final Ligand Database

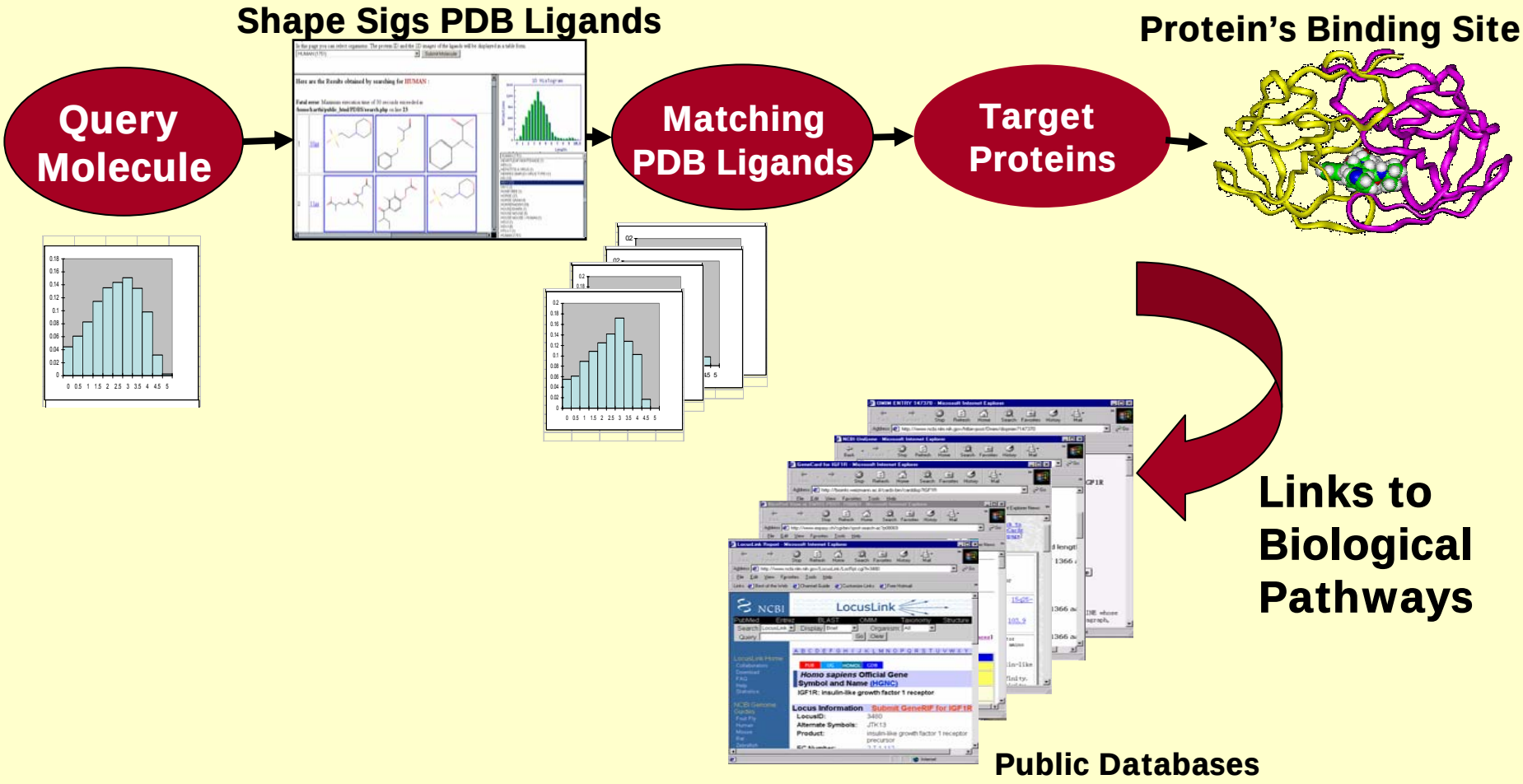
Convert to ShapeSig



PDB-extracted Shape Signature Database

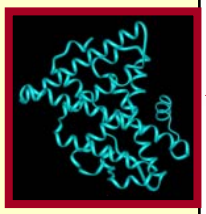

PDB-extracted Shape Signature Database
2D or 3D structure
Query Molecule
List of Hits With high similarity to query
• drug-like features (Lipinski)
• parent protein ID, structure, link
• protein functionality
• pathway info
Potential Toxicities

Molecules → Target Protein → Mechanism



Chemical → Target Protein → Mechanisms

Protein Data Bank (PDB): World Repository of ~35,000
Protein-Ligand Crystal Structures (http://www.rcsb.org/pdb/)



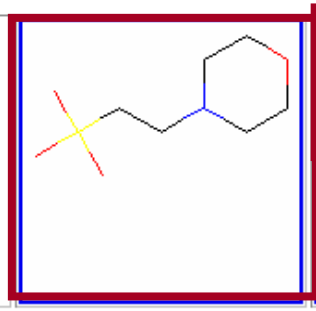
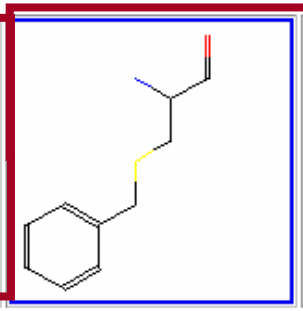
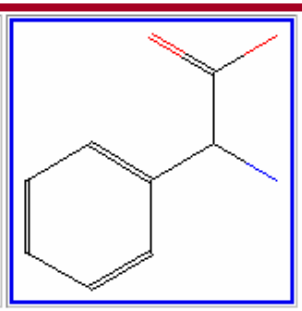
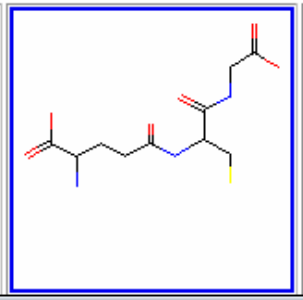
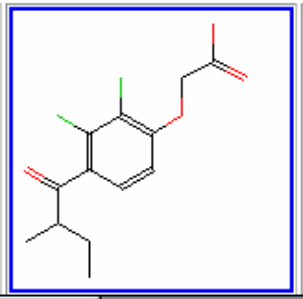
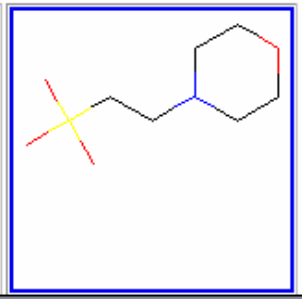
In this page you can select organisms. The protein ID and the 2D images of the ligands will be displayed in a table form.

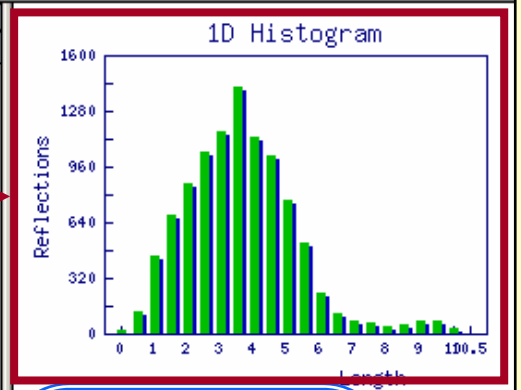
HUMAN (1751) Submit Molecule

Shape Signatures of PDB-extracted ligands

Here are the Results obtained by searching for HUMAN :

← **Protein Structure PDB ref code #**

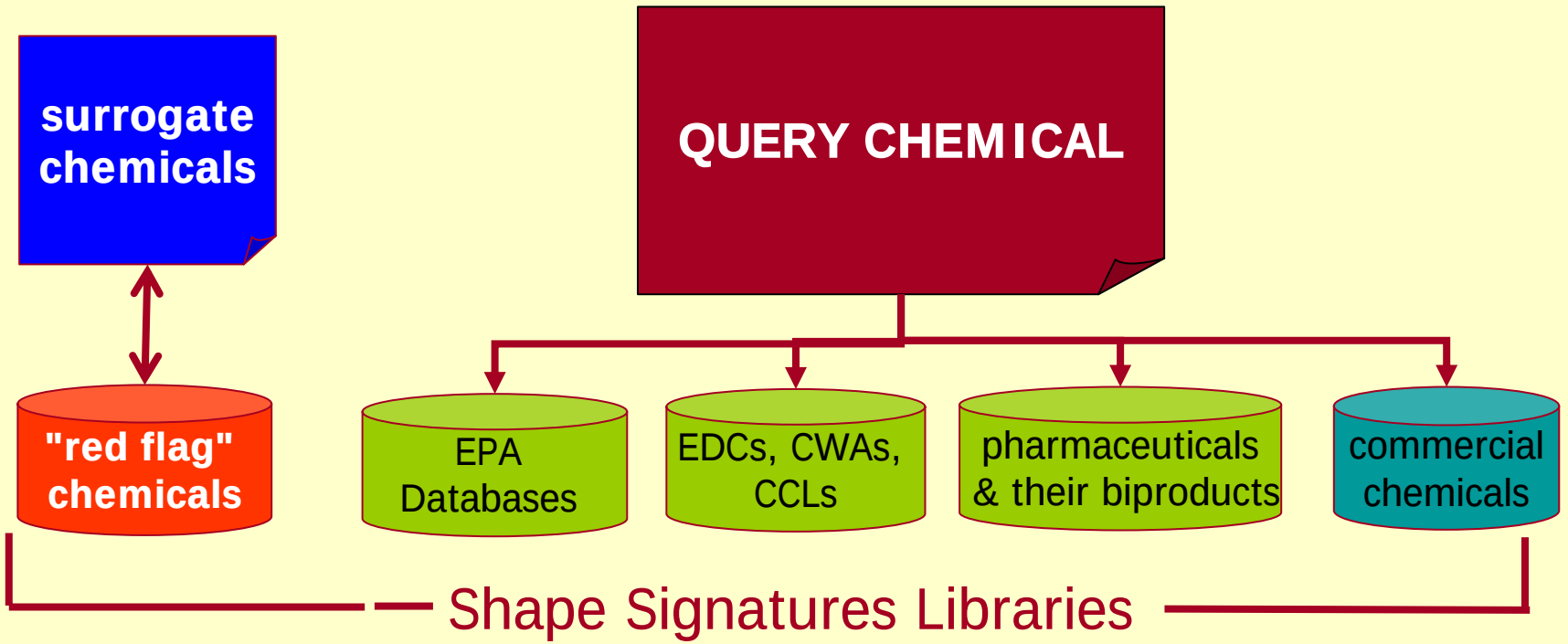
1	10gs			
2	11gs			



HUMAN (1751)
HEARTLEAF NIGHTSHADE (1)
HEN (1)
HEPATITIS A VIRUS (1)
HERPES SIMPLEX VIRUS TYPE-I (1)
HIV (10)
HIV-1 (72)
HIV-2 (3)
HONEYBEE (1)
HORSE (37)
HORSE GRAM (4)
HORSERADISH (28)
HOUNDSHARK (1)
HOUSE MOUSE (5)
HOUSE MOUSE + HUMAN (1)
HSV2 (1)
HSV-I (6)
HTLV-1 (1)
HUMAN (1751)

Species / Protein Family

Identifying Problem Chemicals & Possible Surrogates



Key Features of *Shape Signatures*

- Compact: Encodes molecular shape and other biorelevant features in a single entity
- Finds hits missed by techniques that search on chemical (sub)structure
- User Oriented: fast, simple, intuitive
- No need for special procedures, e.g., molecular alignment, descriptor selection
- No need to reformulate model as more data are added
- Versatile: Works for any number or type of molecular species (organics, organometallics, ions, etc.)
- Complementary to other computational approaches, e.g., QSAR, receptor-based docking/scoring
- Operable stand-alone, or together with other approaches, modules, tools

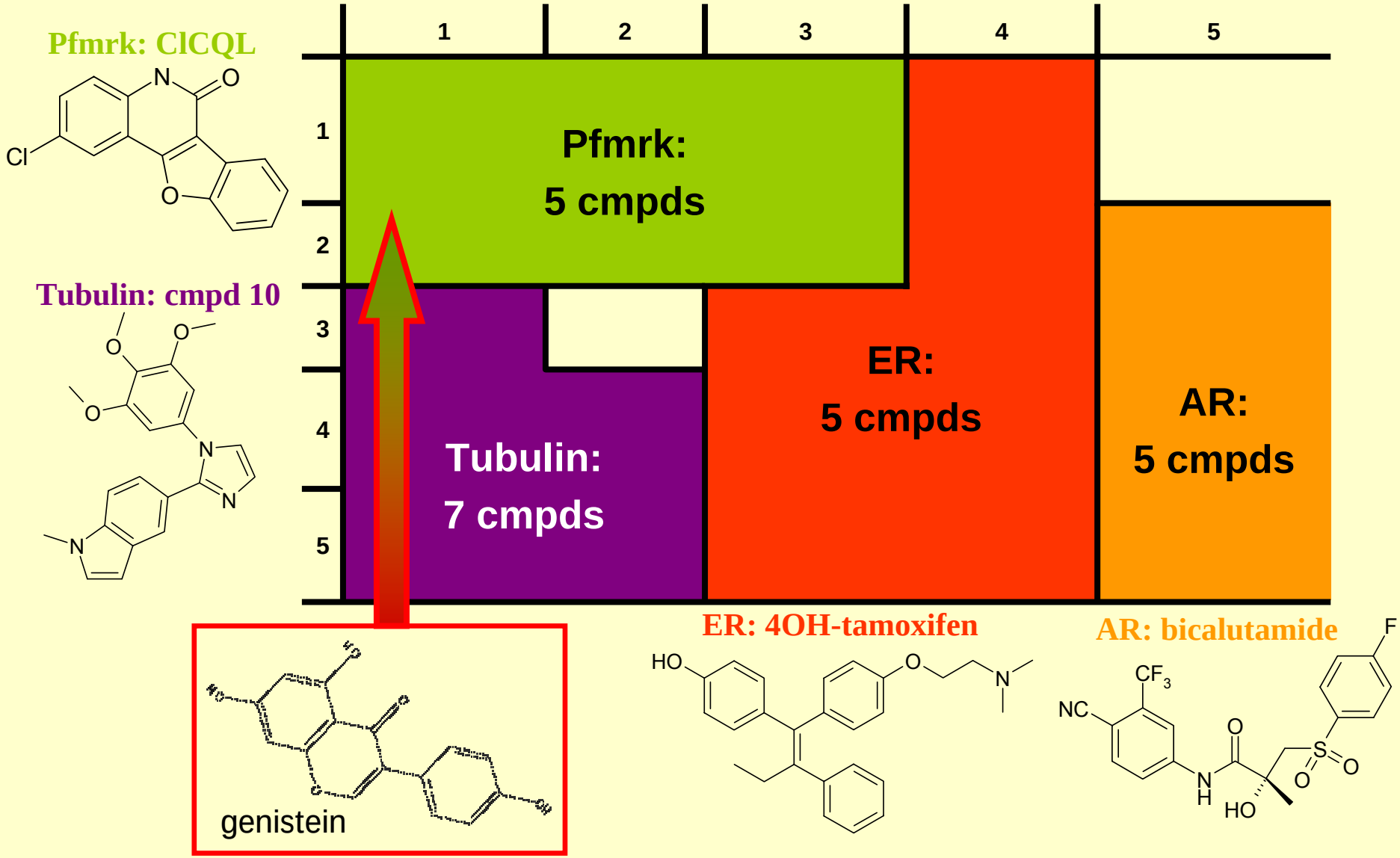
Case Study I

Using Shape Signatures to Cluster Compounds into Functional Groups

Clustering of 22 Bioactive Compounds: AR (5), ER (5), Pfmrk (5) and Tubulin (7) ligands based on Shape Signatures histograms

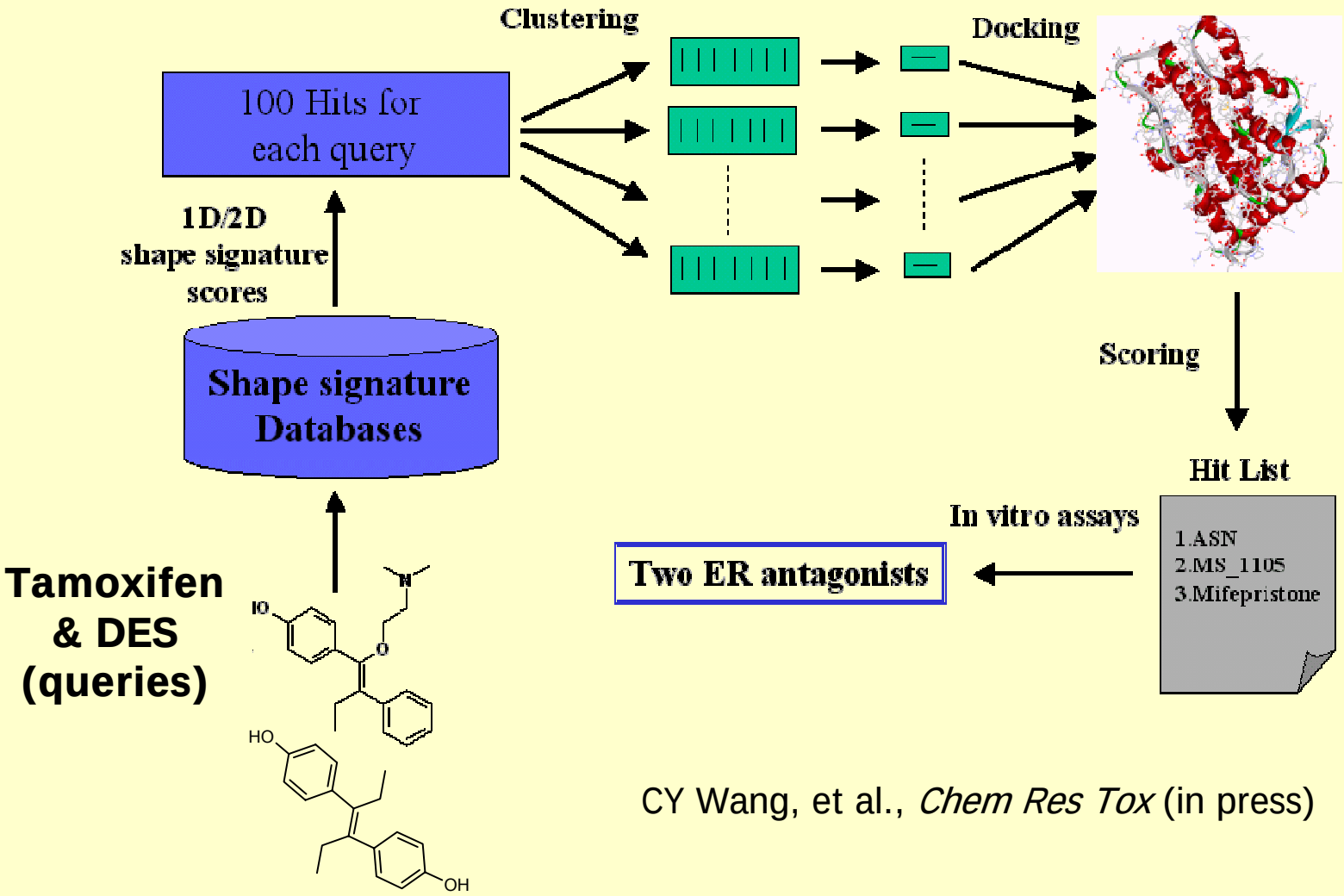
	1	2	3	4	5
1	CICQL, HOBQL, OXINDOLE GENISTEIN			RALOXIFENE	
2	HOPQL		mol28E		BICALUTAMIDE, HOF
3	30b		TAMOXIFEN		
4	25b, 25g		DES		compd18
5	08b, 10, 24h, 40b			ESTRADIOL	DHT, R1881

Method: Kohonen Self-Organizing Map (SOM) **Distance function:** Euclidean


Pfmrk: ClCQL
Cl
N
O
O
Tubulin: cmpd 10
O
O
O
N
N
N
1
2
3
4
5
1
2
3
4
5
Pfmrk:
5 cmpds
Tubulin:
7 cmpds
ER:
5 cmpds
AR:
5 cmpds
genistein
ER: 4OH-tamoxifen
HO
O
N
AR: bicalutamide
F
CF3
NC
O
N
H
HO
S
O
O

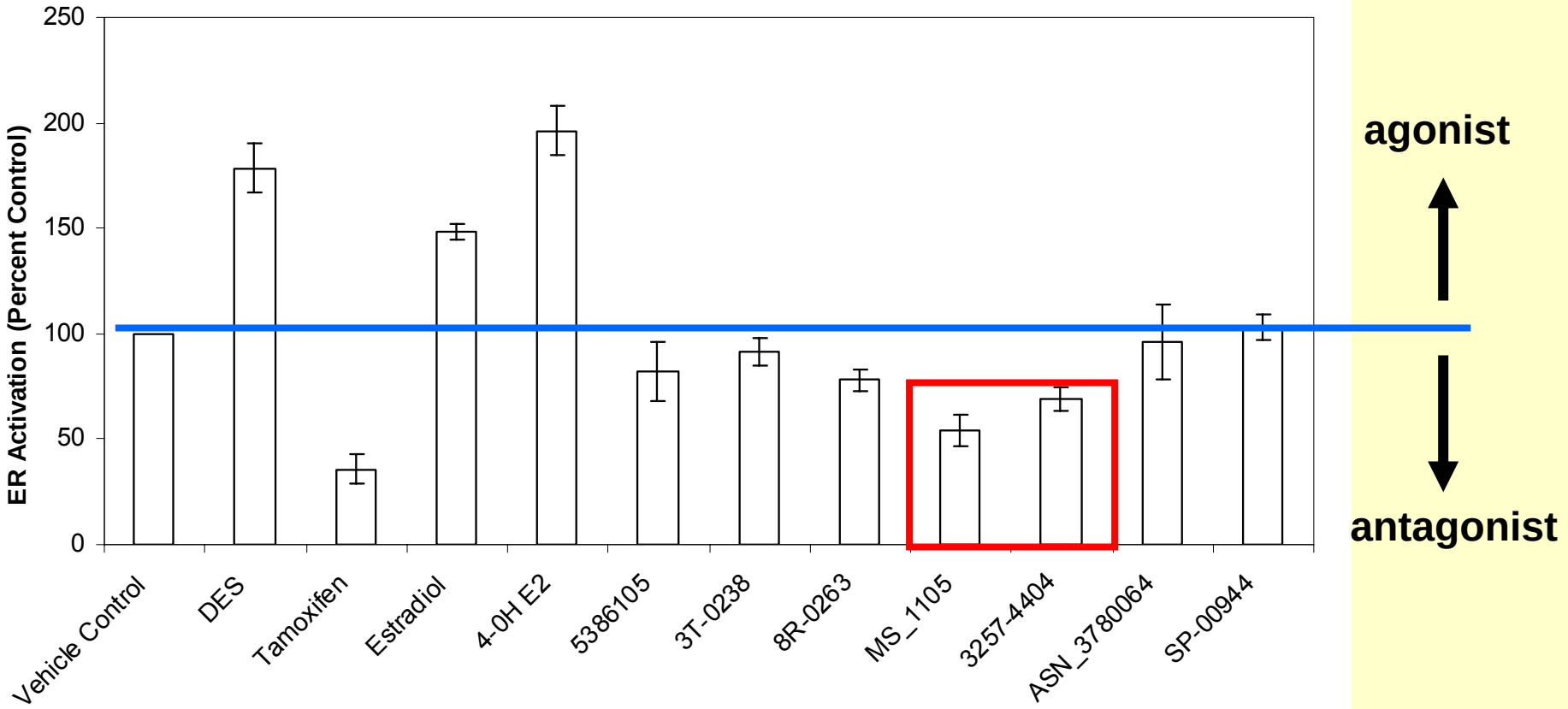
Case Study II

Identification of Previously Unrecognized (Anti)estrogens

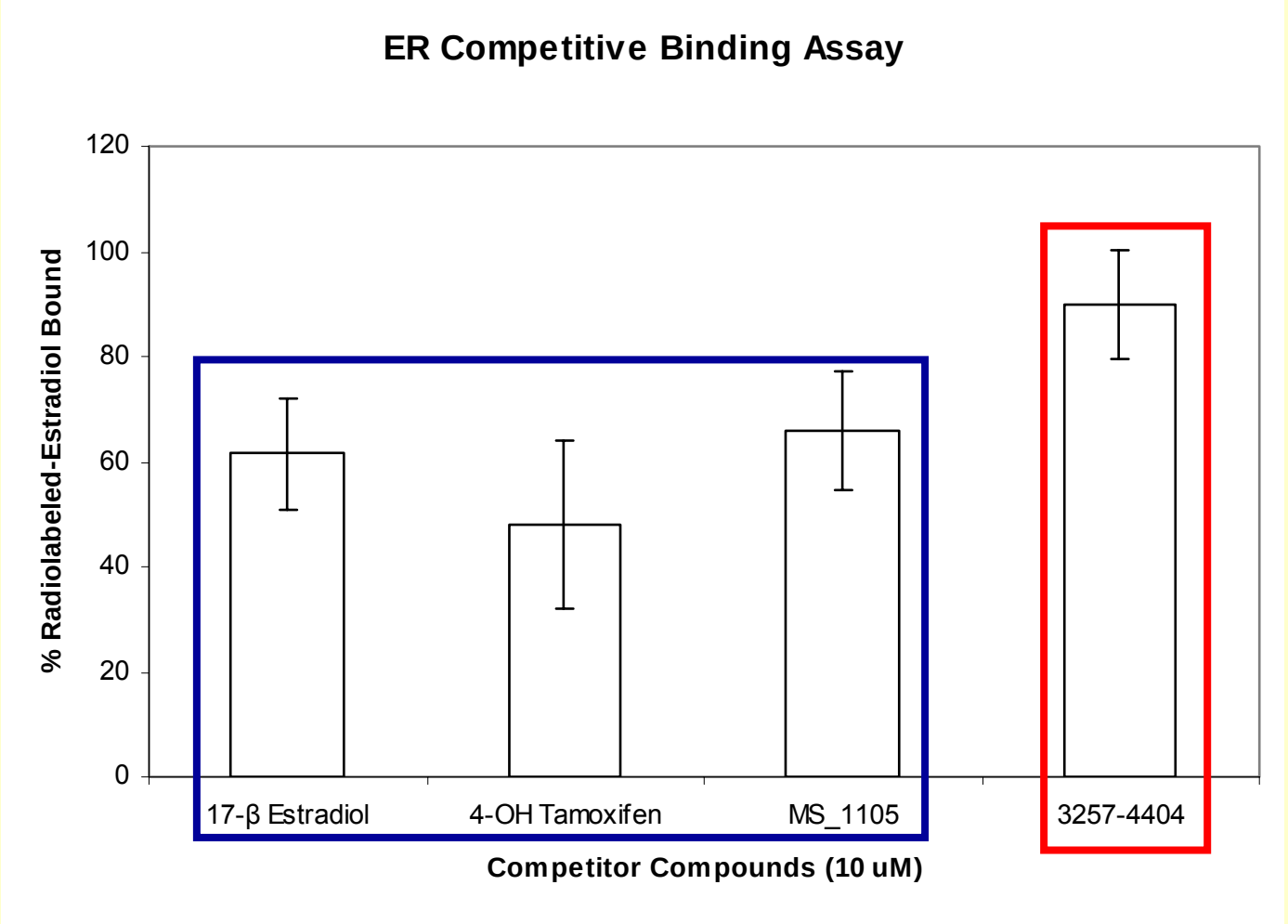


CY Wang, et al., *Chem Res Tox* (in press)

Estrogenic Activity Assay:
Identify ER agonists and antagonists



ER Competitive Binding Assay

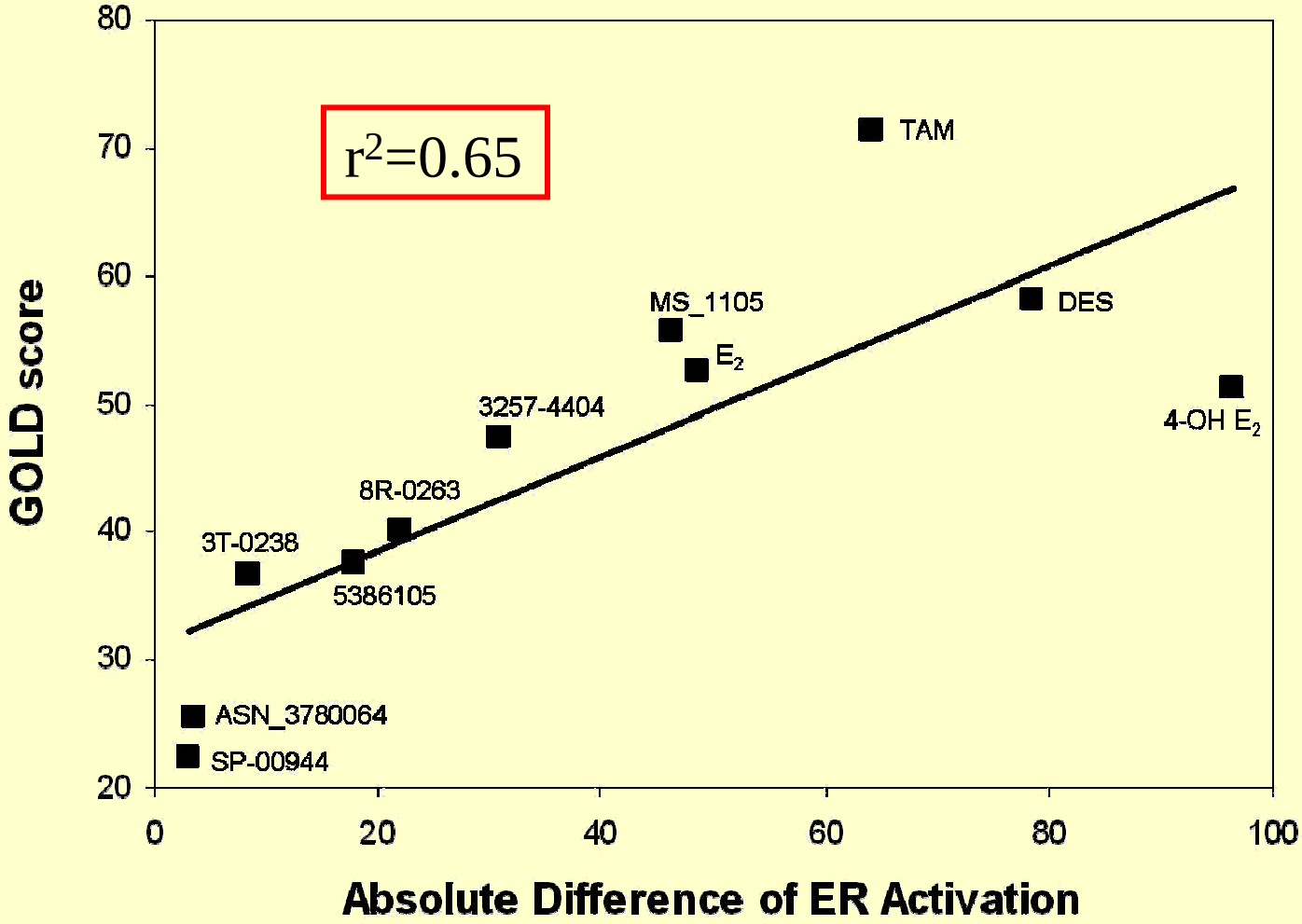


MS_1105 competes with radiolabeled estradiol

3257-4404 has lower affinity to ER

Both compounds are ligands of ER

Correlation of GOLD docking score and absolute difference of ER activation

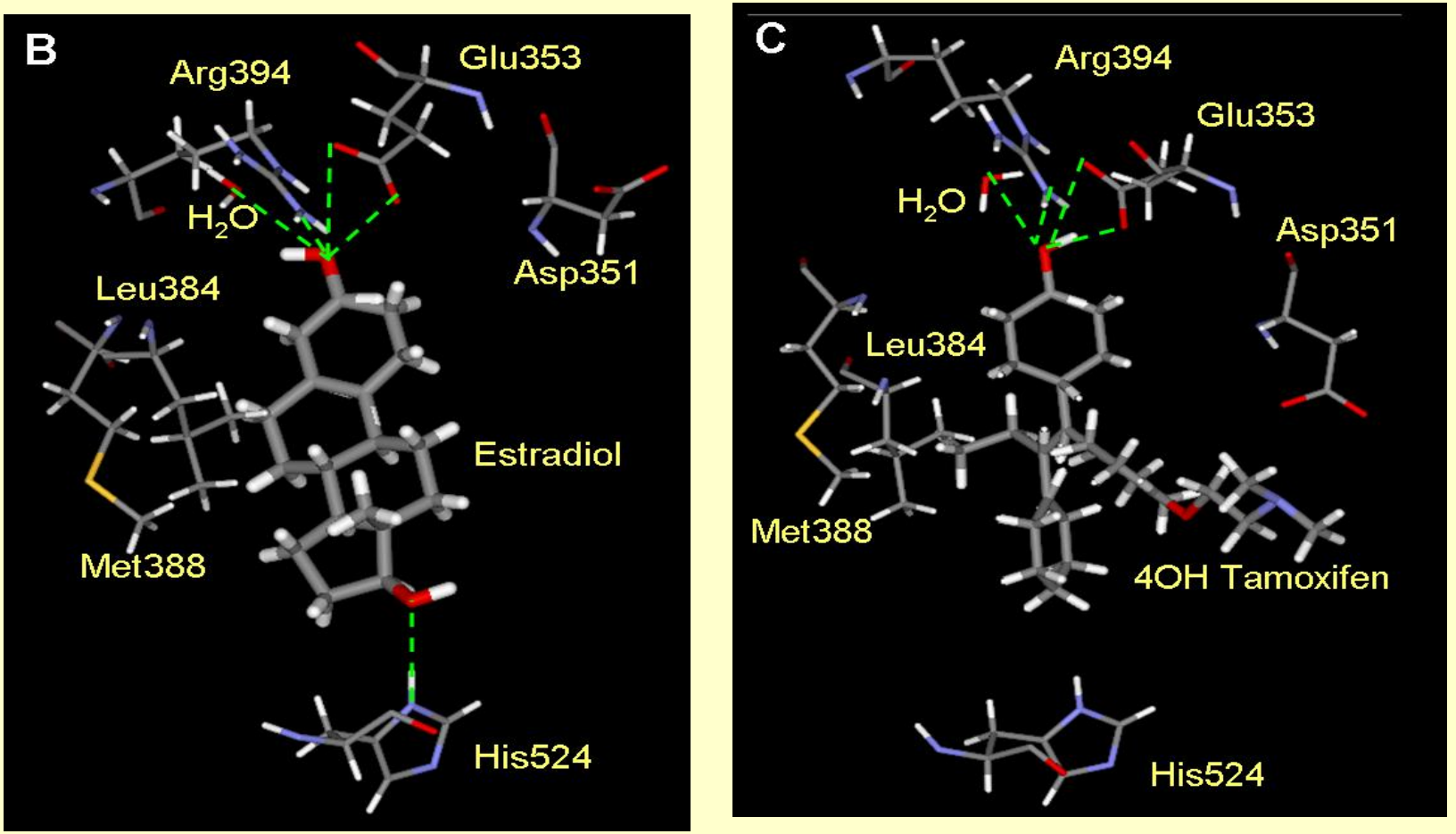


Molecular Modeling of ER-ligand complex

Green: 4OH-Tamoxifen
Red: Raloxifene
Blue: MS_1105
Yellow: 3257-4404

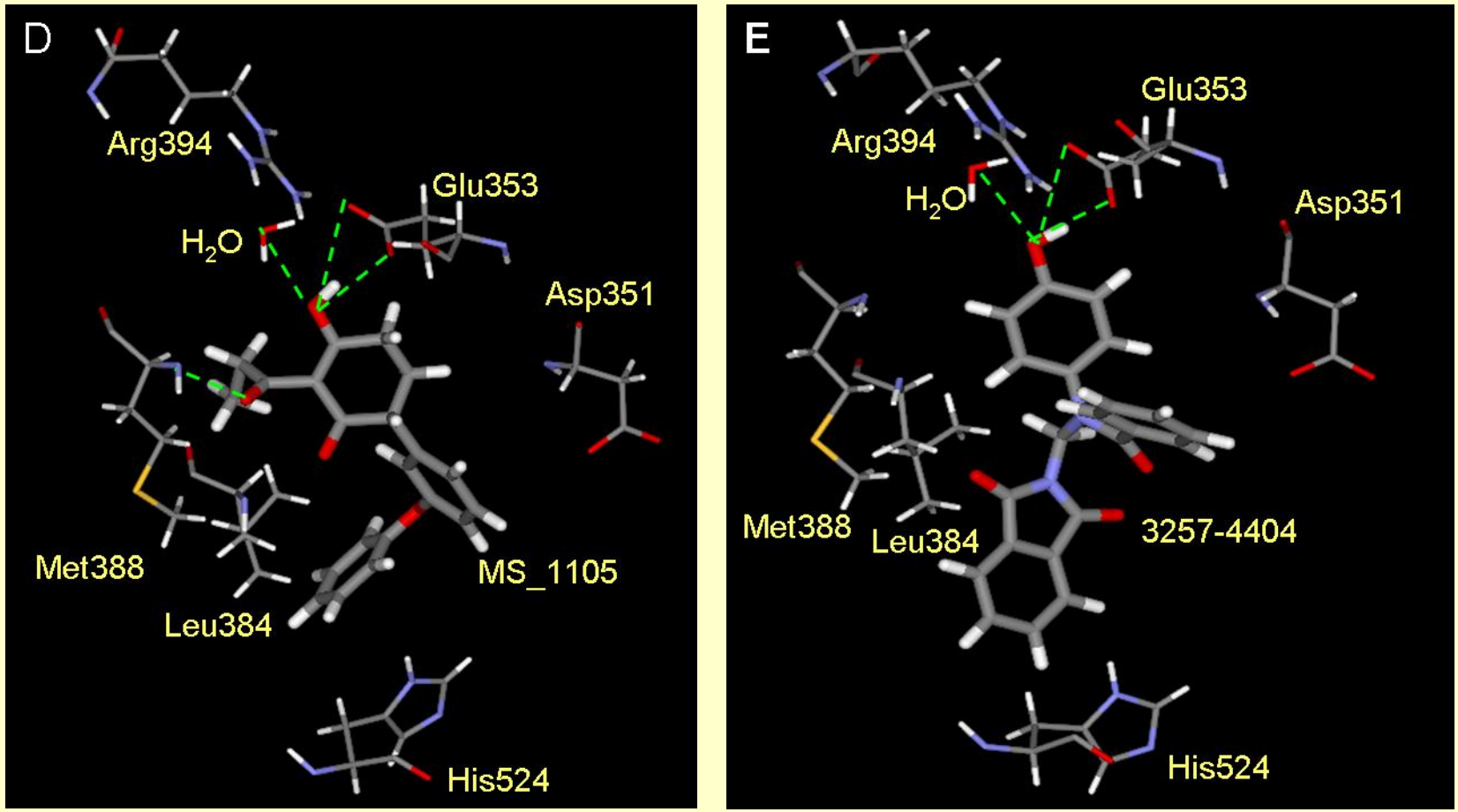
ER docking orientation for MS_1105 and 3257-4404 is similar to 4-OH tamoxifen and raloxifene

Hydrogen bonding network



Hydrogen bond

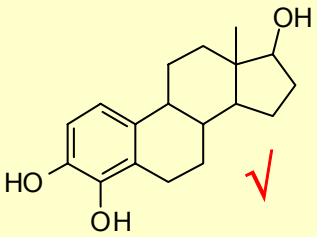
Hydrogen bonding network – Cont'd



Hydrogen bond

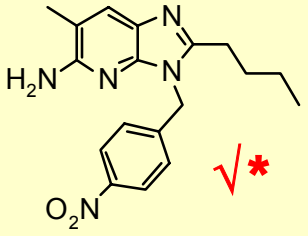
Predicted to Exhibit (Anti-)estrogenic Activity

Vendor	Compound
ASINEX	5386105, ASN_3780064
BIONET	3T_0238, 8R_0263, MS_1105
Sigma	4-OH Estradiol
Maybridge	SP-00944
ChemDiv	3257-4404



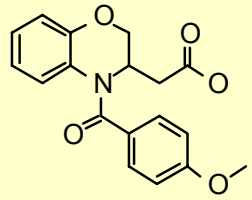
√

4-OH Estradiol



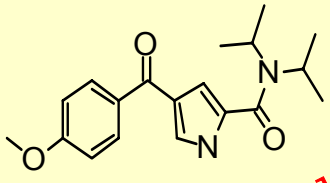
√*

5386105



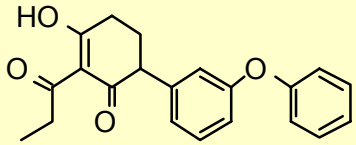
√*

3T_0238



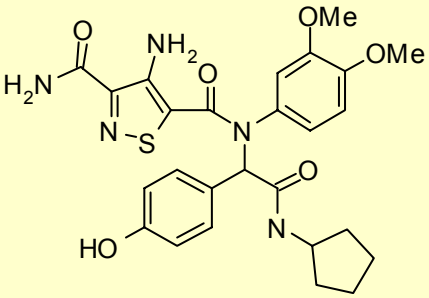
√*

8R-0263

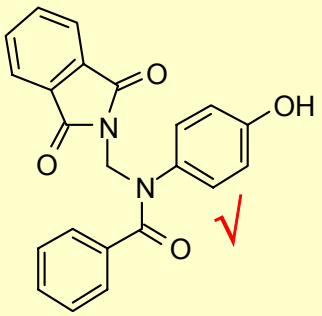


√

MS_1105

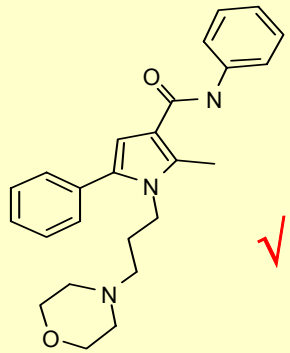


ASN_3780064



√

3257-4404



√*

SP-00944

Case Study III

Virtual Screening of Estrogen Receptor Ligands

- Enrichment study on 21 estrogenic compounds in the PDB-extracted shape signature database.
- *Using each ER ligand once as query, can it find the other 20 ER ligands?*
- Results evaluated based on the enrichment score (E), which is defined as the ratio of yield of actives in the hitlist (*Ha*/*Ht*) relative to the yield of actives in the database (*A*/*D*)

$$E = \frac{\frac{H_a}{H_t}}{\frac{A}{D}}$$

H_a = number of estrogenic compounds in top 20 hit list
H_t = 20
A =21
D = 5432

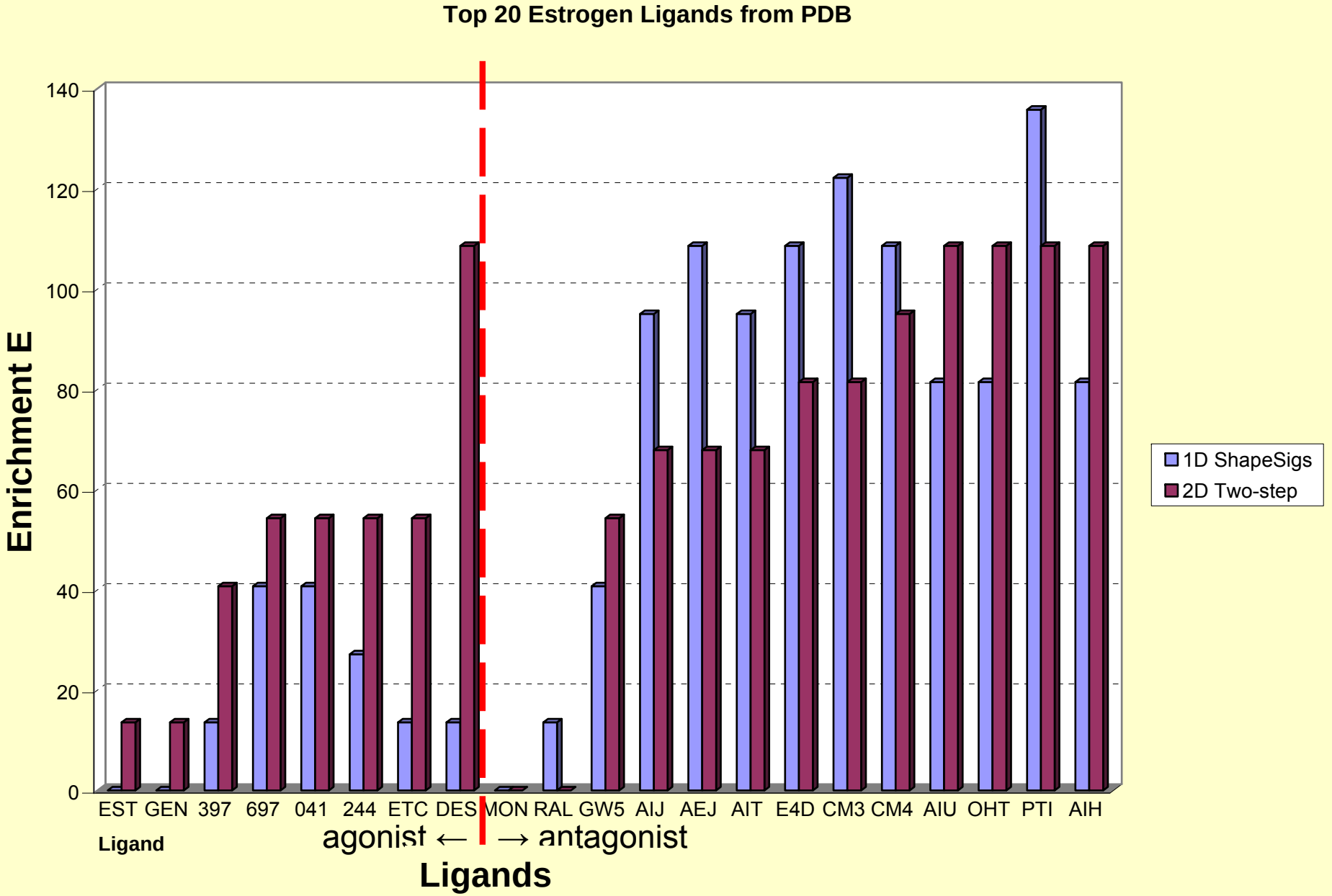
Theoretical perfect E = 5432/21 = 258

E = 1, the virtual screening method is no better than random sampling

E = 5, a 5-fold improvement over random selection.

E >10, very good

E > 100, exceptional and extremely rare.


Top 20 Estrogen Ligands from PDB
Enrichment E
140
120
100
80
60
40
20
0
EST GEN 397 697 041 244 ETC DES MON RAL GW5 AIJ AEJ AIT E4D CM3 CM4 AIU OHT PTI AIH
Ligand
agonist ← → antagonist
Ligands
1D ShapeSigs
2D Two-step

Possible Ways Shape Signatures May Be Useful to EPA Investigators

- **Problem:** What is possible target or mode of action for my compound(s)?
- **Approach:** Use compound as query to search PDB-based ShapeSigs database
- **Problem:** Are there alternative protein targets (intended or unintended) for my compound(s)?
- **Approach:** Use compound as query, and explore list of protein 'hits' in PDB-based ShapeSigs database
- **Problem:** Is a certain compound active or inactive (e.g., ER ligand)?
- **Approach:** Compare ShapeSig of compound vs. ShapeSig database of known actives

Possible Ways Shape Signatures May Be Useful to EPA Investigators (cont ..)

- **Problem:** Is it possible to tell whether a collection of compounds might have common modes of action?
- **Approach:** Cluster compounds into groups based on their ShapeSigs
- **Problem:** Are there compounds in a database that have similar endpoints (e.g., physical props, functionality, mode of action) to my compound(s)?
- **Approach:** Use known compounds as queries to search database (virtual screening), using ShapeSigs to find matches

Thank You!

welshwj@umdnj.edu