ebCTC environmental bioinformatics and Computational Toxicology Center

www.ebCTC.org

Biological Spectra Analysis of the ToxCast Chemicals: Linking Bioactivity Profiles to Molecular Structure

William Welsh

D Chekmarev, V Kholodovych, N Ai, W Tong (FDA), P Georgopoulos

UMDNJ-Robert Wood Johnson Medical School, Piscataway, NJ

welshwj@umdnj.edu

Funded by the USEPA Science to Achieve Results (STAR) Grant RD-83272101

Biological spectra analysis (BSA):

Link biological activity profiles to molecular structures

- Traditional (Q)SAR methods use the structure-based features (molecular descriptors) of a collection of chemicals to describe and compare their biological activities.

molecular structure ⇒ bioactivity

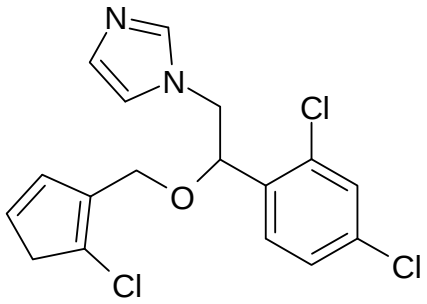
- In contrast, BSA uses the biological response profiles of the chemicals to describe and compare their molecular structures.

molecular structure ⇐ bioactivity

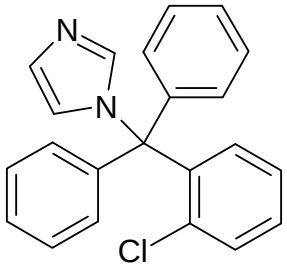
Fliri AF et al PNAS 2005, 12(2), 261-266

BSA: A Simple Example

molecular structure ⟸ bioactivity



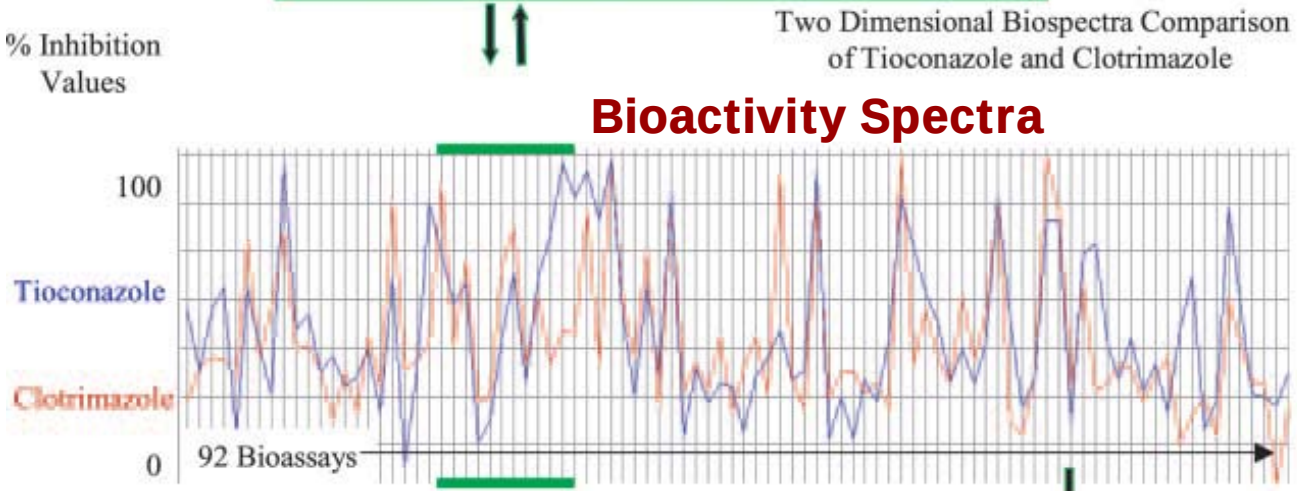
Tioconazole



Clotrimazole

	Ca channel (DHP)	Ca channel (VERA)	CB1	CCKA	CCR1	chol trans	Cl channel	COMT	COX2	CYP 1a2	CYP 2b6	CYP 2C19
Clotrimazole	88	21	56	-3	0	53	70	13	41	13	27	26
Tioconazole	59	38	47	-19	-10	23	51	6	49	67	97	82

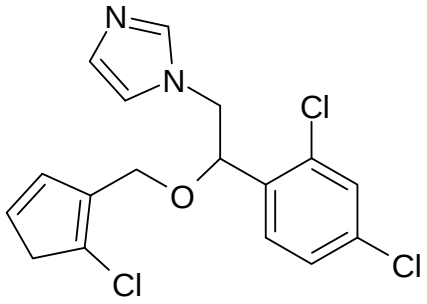
% Inhibition Values (10 µM)



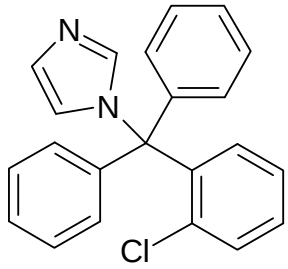
Fliri AF et al PNAS 2005, 12(2), 261-266

Biological Activity Spectra

- depicted as a Heat Map -



Tioconazole



Clotrimazole

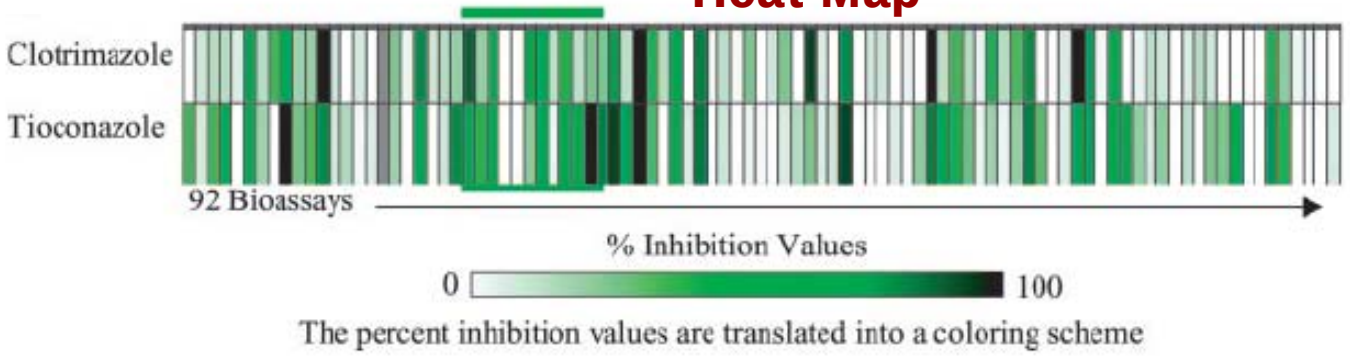
	Ca channel (DHP)	Ca channel (VERA)	CB1	CCKA	CCR1	chol trans	Cl channel	COMT	COX2	CYP 1a2	CYP 2b6	CYP 2C19
Clotrimazole	88	21	56	-3	0	53	70	13	41	13	27	26
Tioconazole	59	38	47	-19	-10	23	51	6	49	67	97	82

% Inhibition Values (10 µM)

% Inhibition Values

Heat Map
One Dimensional Biospectra Comparison of Tioconazole and Clotrimazole

Heat Map



The percent inhibition values are translated into a coloring scheme

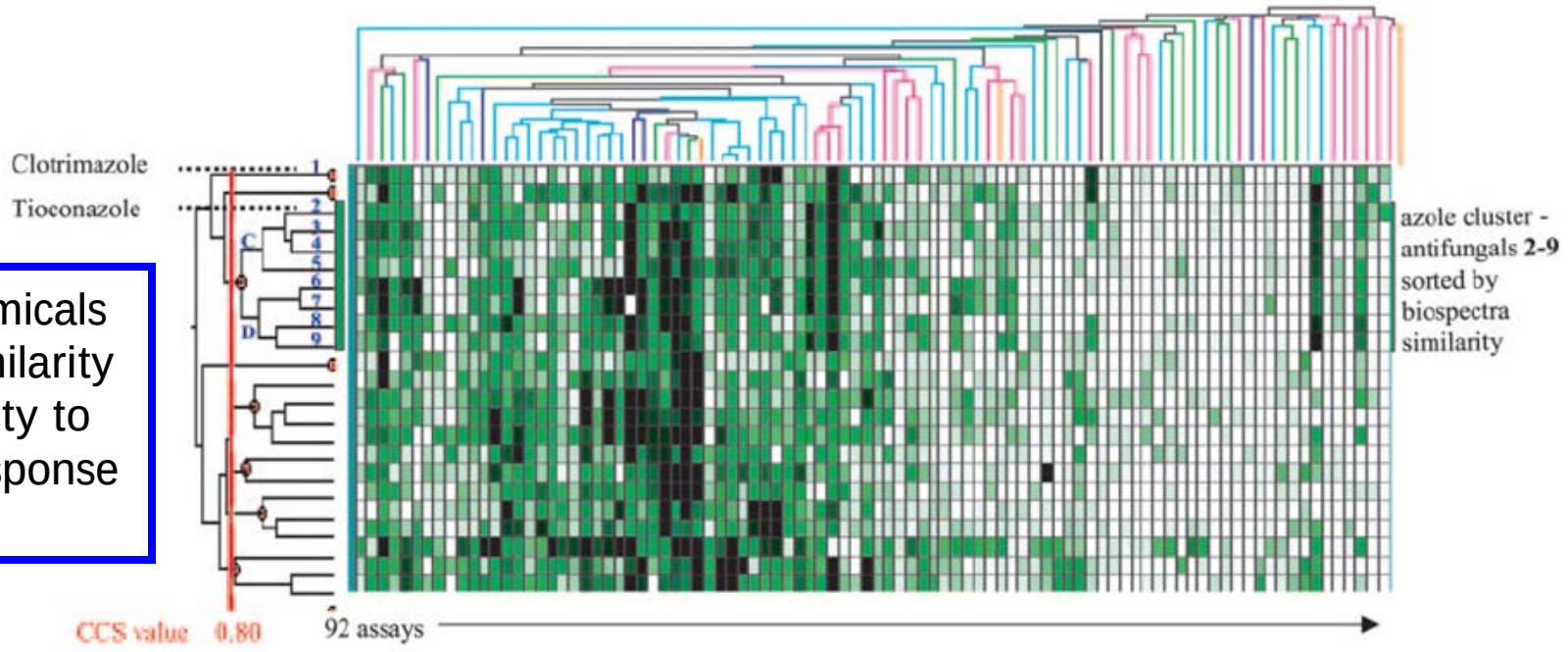
Fliri AF et al PNAS 2005, 12(2), 261-266

Heat Map for a Collection of Chemicals and a Panel of Protein Receptors

Two-way Hierarchical Clustering

clusters proteins based on similarity in their bioresponse profile

clusters chemicals based on similarity in their ability to induce bioresponse profile



Fliri AF et al PNAS 2005, 12(2), 261-266

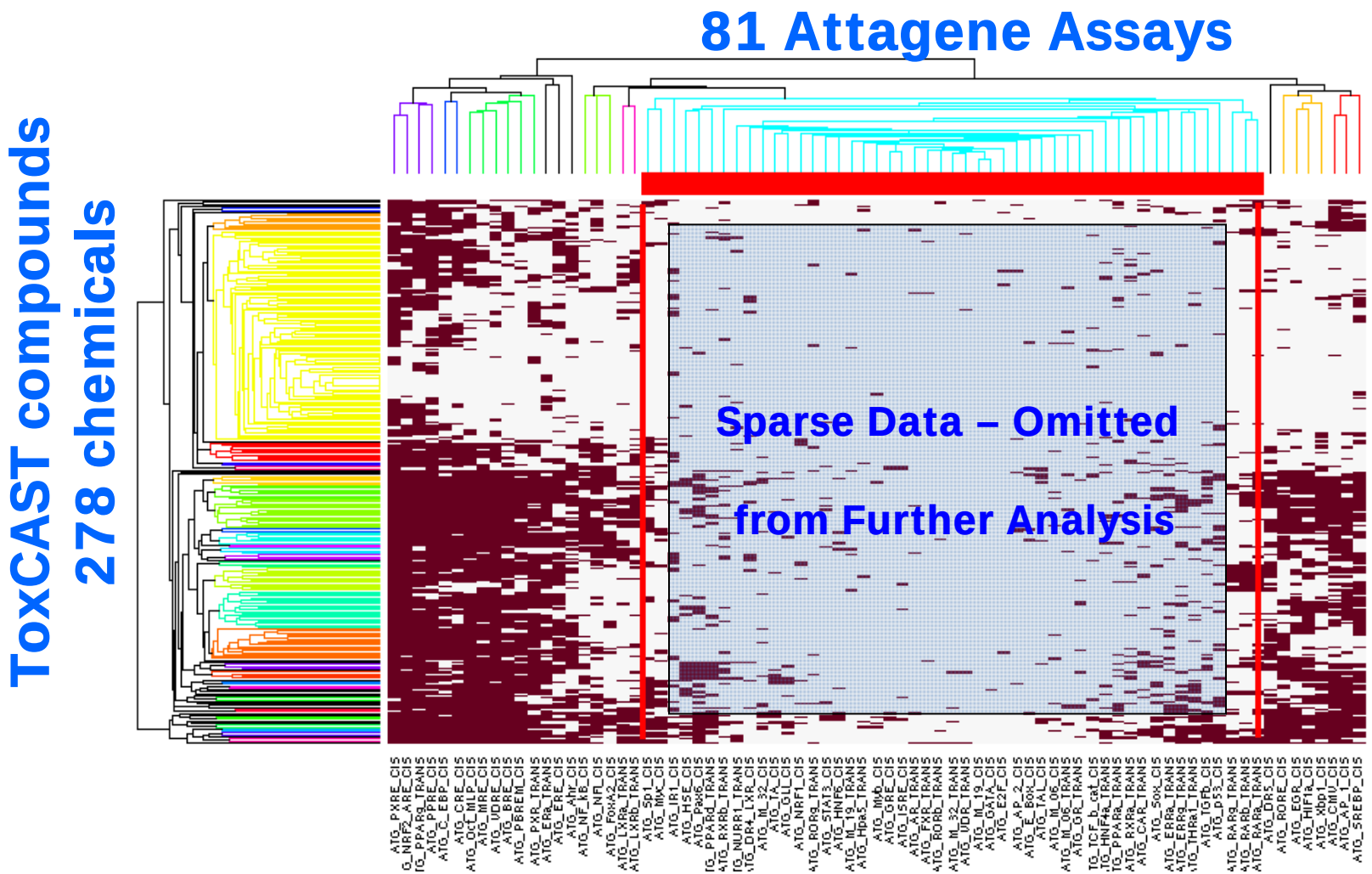
Present Study

BSA Study on Assay Data from Attagene, Inc.

- Transcription Activation (TA) assays
- 309 ToxCast chemicals @ 81 assays
- Reported LEL (lowest effective level) values from each assay
- *Inactive* chemical-assay combinations were assigned LEL =1000000

- Two-way hierarchical (UPGMA) clustering was run using Bioinformatics Toolbox v.3.1, MATLAB 7.6
- Analysis employed both Euclidean distance and Cosine metrics
- Assay results and calculated molecular descriptors were pre-processed using Unsupervised Forward Selection (UFS)
  - removes co-variant, redundant, and non-varying descriptors

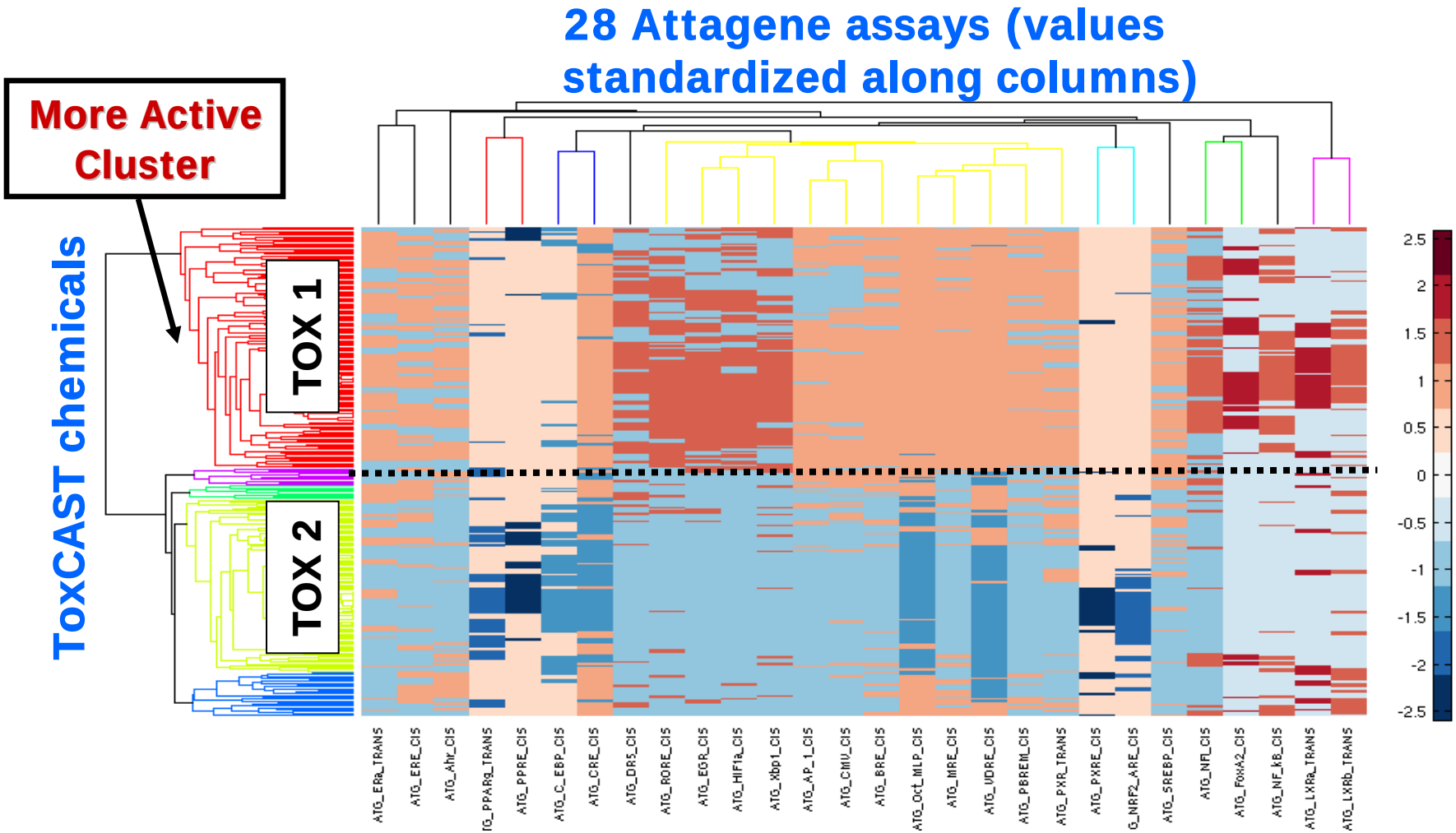
BSA Study on Attagene Data



Dark regions - Compounds with measurable reported LEL (lowest effective level).

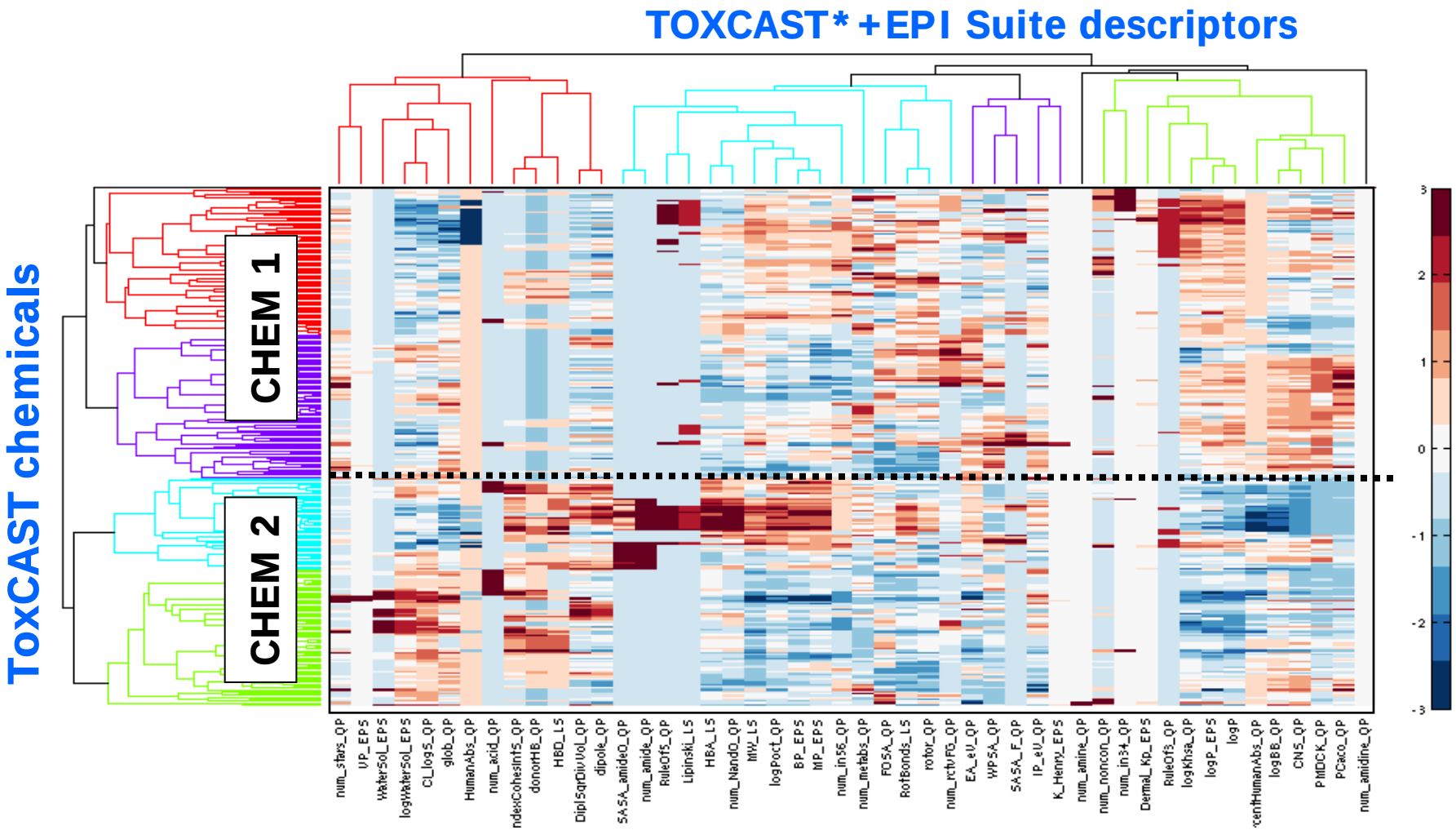
White regions- *Inactive* chemical-assay combinations (LEL =1000000).

Heat Map for Reduced Set of 28 Assays



Hierarchical clustering of ToxCast chemicals in the space of Attagene response biospectra reveals two major clusters **TOX1** (red) and **TOX2** (violet-green-lime-blue).

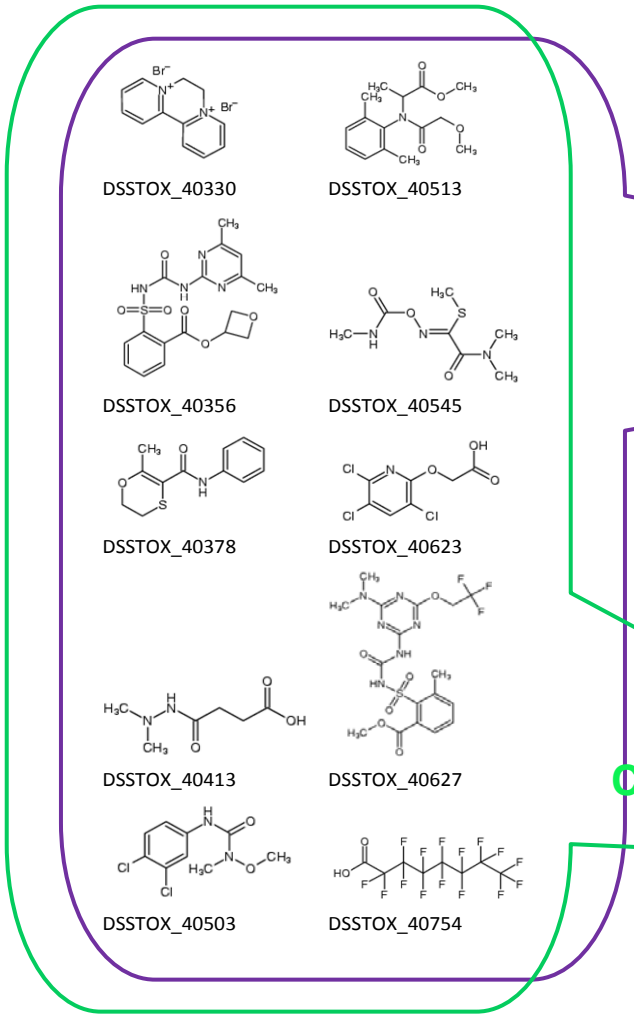
Heat Map for Space of Chemical Descriptors



Hierarchical clustering of ToxCast chemicals in chemical descriptor space reveals two major clusters **CHEM1** (red-violet) and **CHEM2** (blue-green).

* TOXCAST: a combined set of Leadscope, QikProp and PhysChem derived descriptors

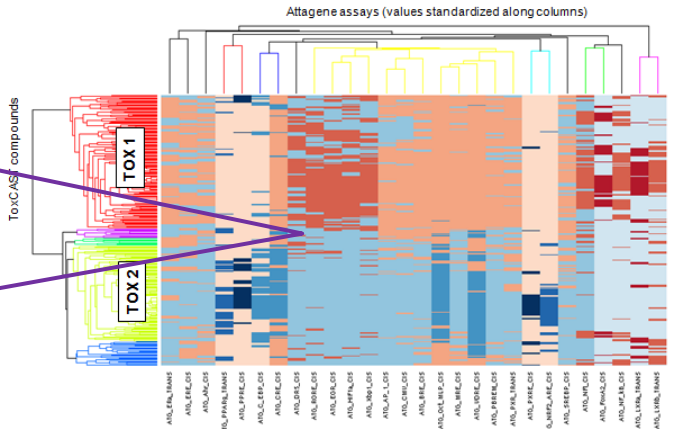
Connection between Similarities in Biospectra and Chemical Space



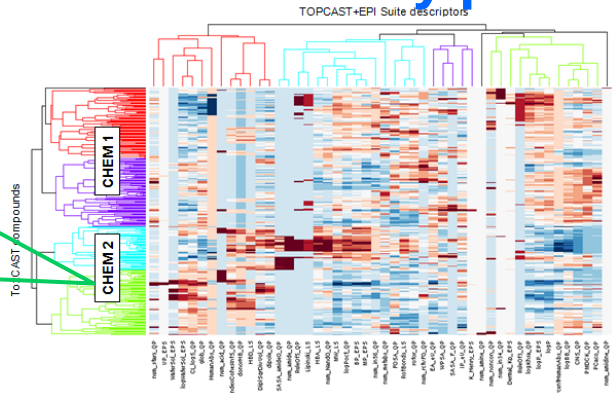
TOX2

CHEM2

TOX similarity profile



CHEM similarity profile



No obvious chemical similarities within individual subclusters.

Cross-mapping of TOX and CHEM Spaces

	CHEM 1	CHEM 2
TOX 1	101 (74%)	36 (26%)
TOX 2	55 (39%)	86 (61%)

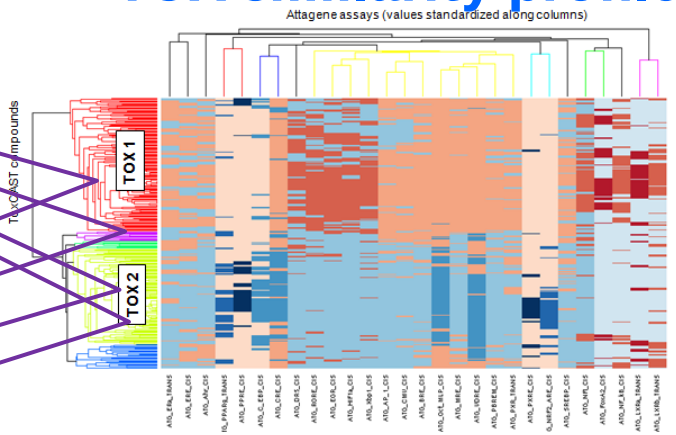
74% of cmpds from cluster TOX1 fall into CHEM1 and 61% of cmpds from TOX2 fall into CHEM2.

Overall association between major clusters of two spaces is found to be 67%.

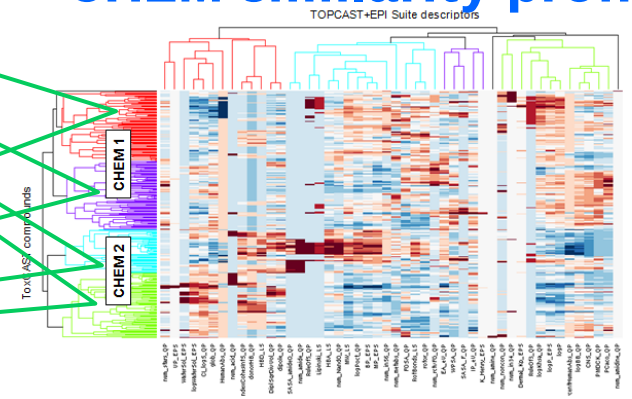
TOX

CHEM

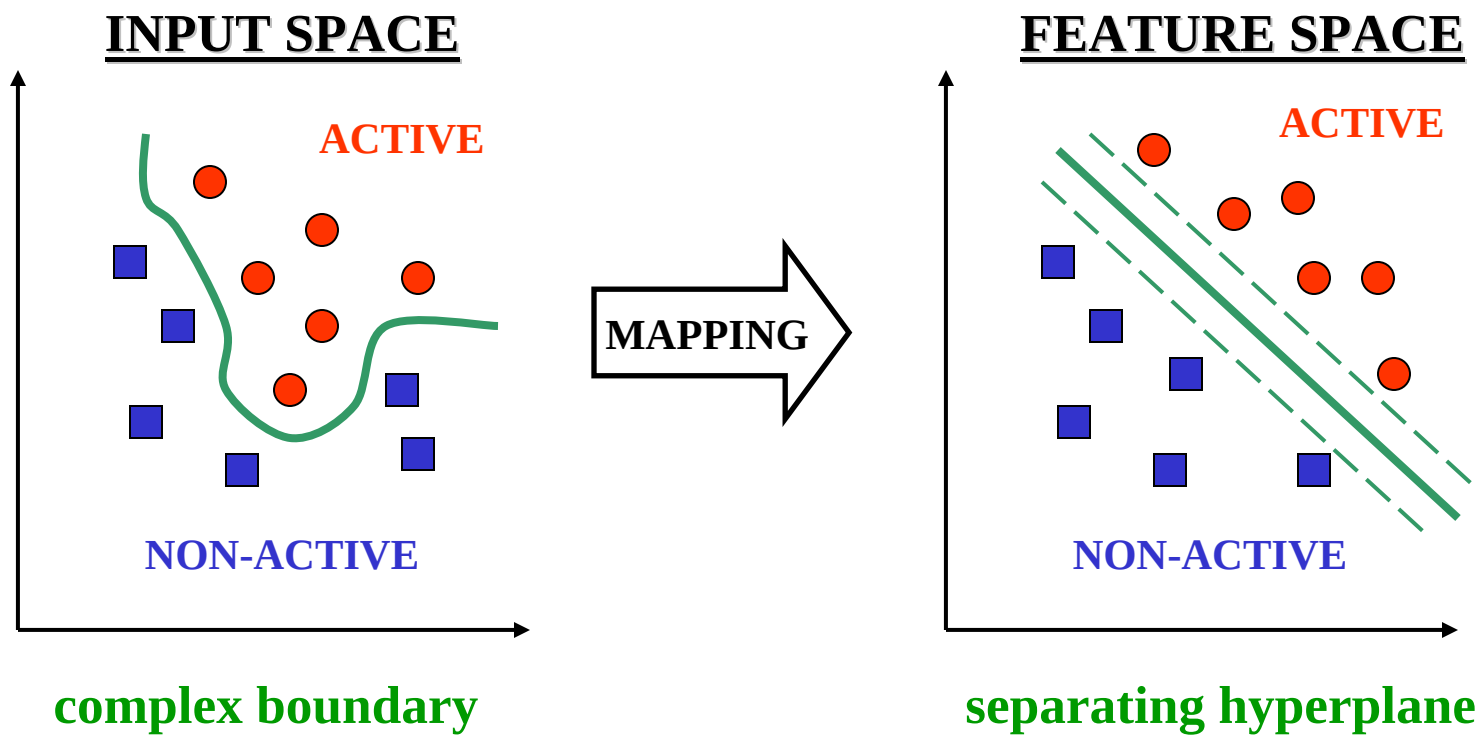
TOX similarity profile



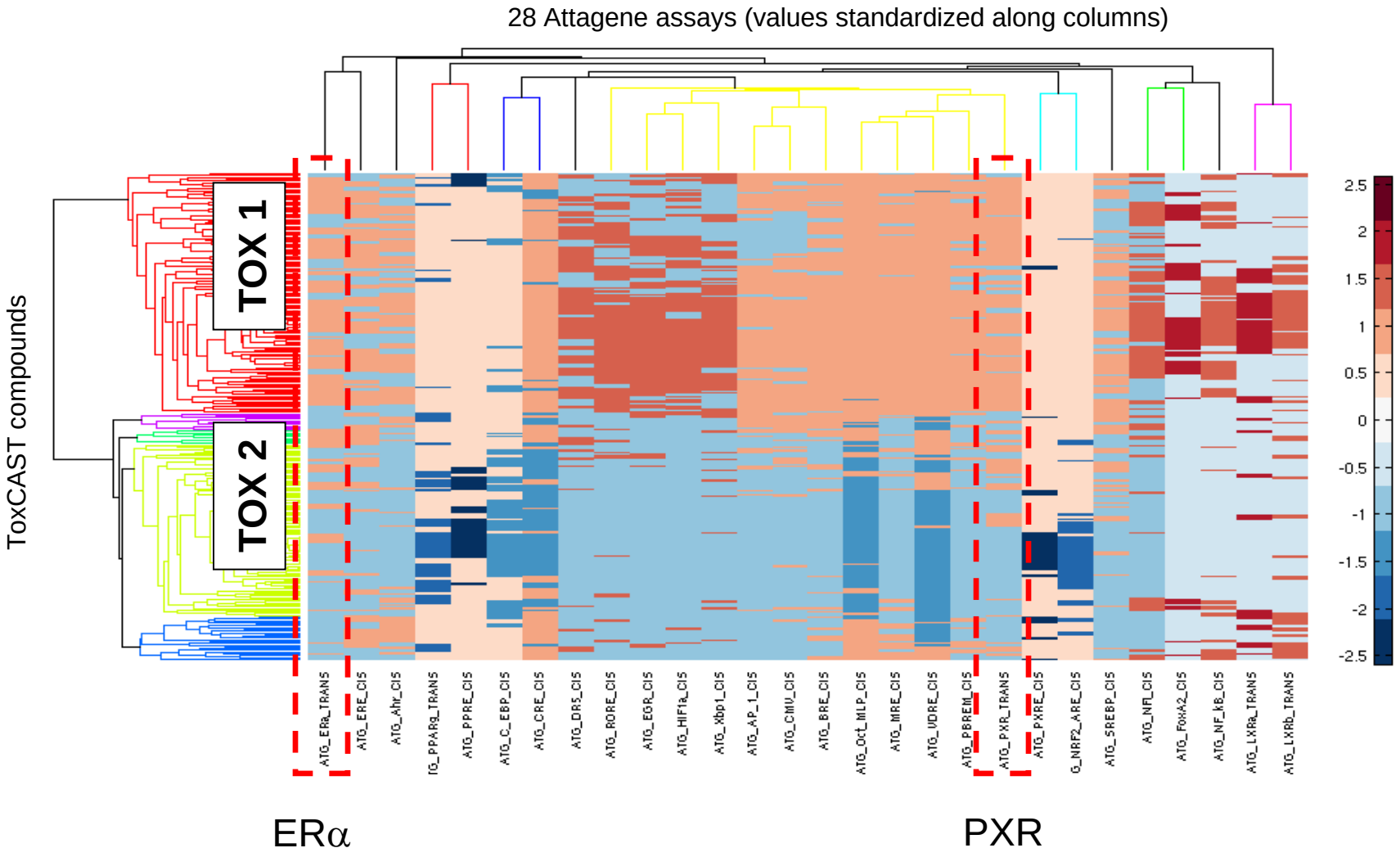
CHEM similarity profile



Classification by Support Vector Machine (SVM)



SVM classification for Two Receptors from the Transcription Factor Assays of Attagene



Support Vector Classification (SVC)

ERα
(active/non-active 50/50% split)

Type of Molecular Descriptors	Number of Molecular Descriptors	Q_{10CV} (%)	AUC ROC
TOXCAST[1]	40	**77**	0.80
TOXCAST + EPISuite	48	**78**	0.83
1DSS[2]	24	**71**	0.70
2DSS[2]	137	**72**	0.74
MOE2D[3]	93	**74**	0.78
1DSS+MOE2D	113	**73**	0.79

PXR
(active/non-active 57/43% split)

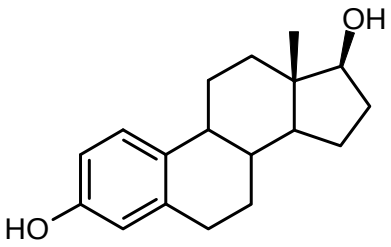
Type of Molecular Descriptors	Number of Molecular Descriptors	Q_{10CV} (%)	AUC ROC
TOXCAST[1]	40	**73**	0.75
TOXCAST + EPISuite	48	**75**	0.76
1DSS[2]	24	**72**	0.70
2DSS[2]	137	**73**	0.75
MOE2D[3]	93	**74**	0.79
1DSS+MOE2D	113	**74**	0.76

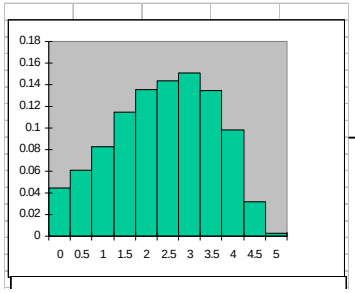
Similar Range of Statistical Quality for Various Descriptor Sets

Shape Signatures

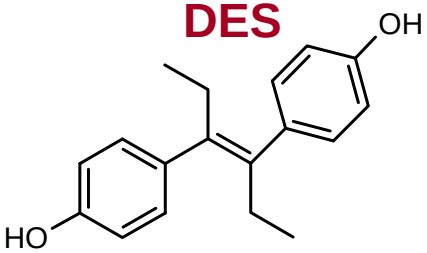
molecules are compared by subtracting their histograms

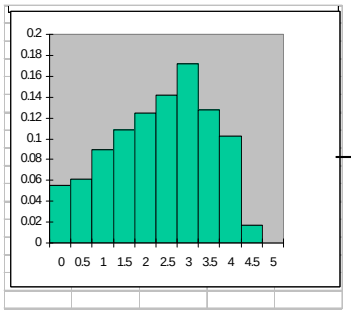
17β-estradiol





DES

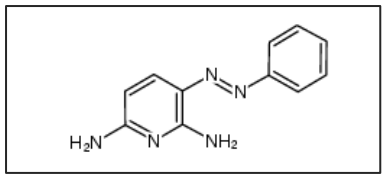




***Diff* = 0.082**

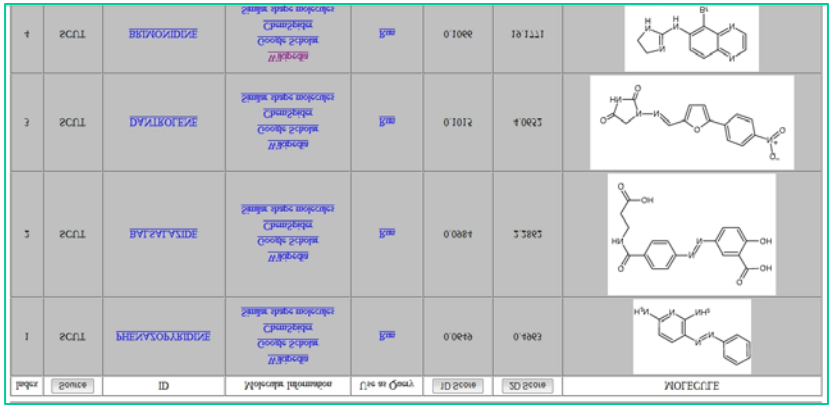
Small *Diff* value means that two molecules have similar shape

Query

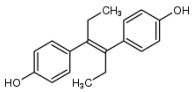
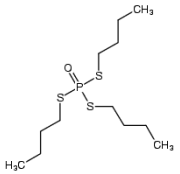
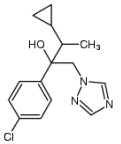
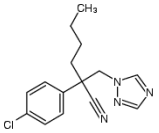
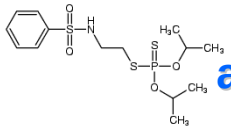
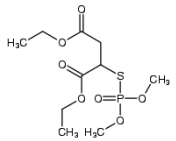


Shape Signatures
Chemical Libraries

Ranked Hit List



Shape Signatures-predicted Structural Analogs of ERα agonist

ToxCast ID	Structure	ER Activity Attagene	Shape Sigs Score X10 *	Gold Score
	 DES	Known agonist	-	-
40510	 Tribufos	33 **active**	0.27	53
40409	 Cyproconazole	33 **active**	0.32	49
40535	 Myclobutanil	100 **active**	0.33	47
40353	 Bensulide	5.82 **active**	0.34	65
40504	 Malaoxon	1000000 **inactive**	0.36	55

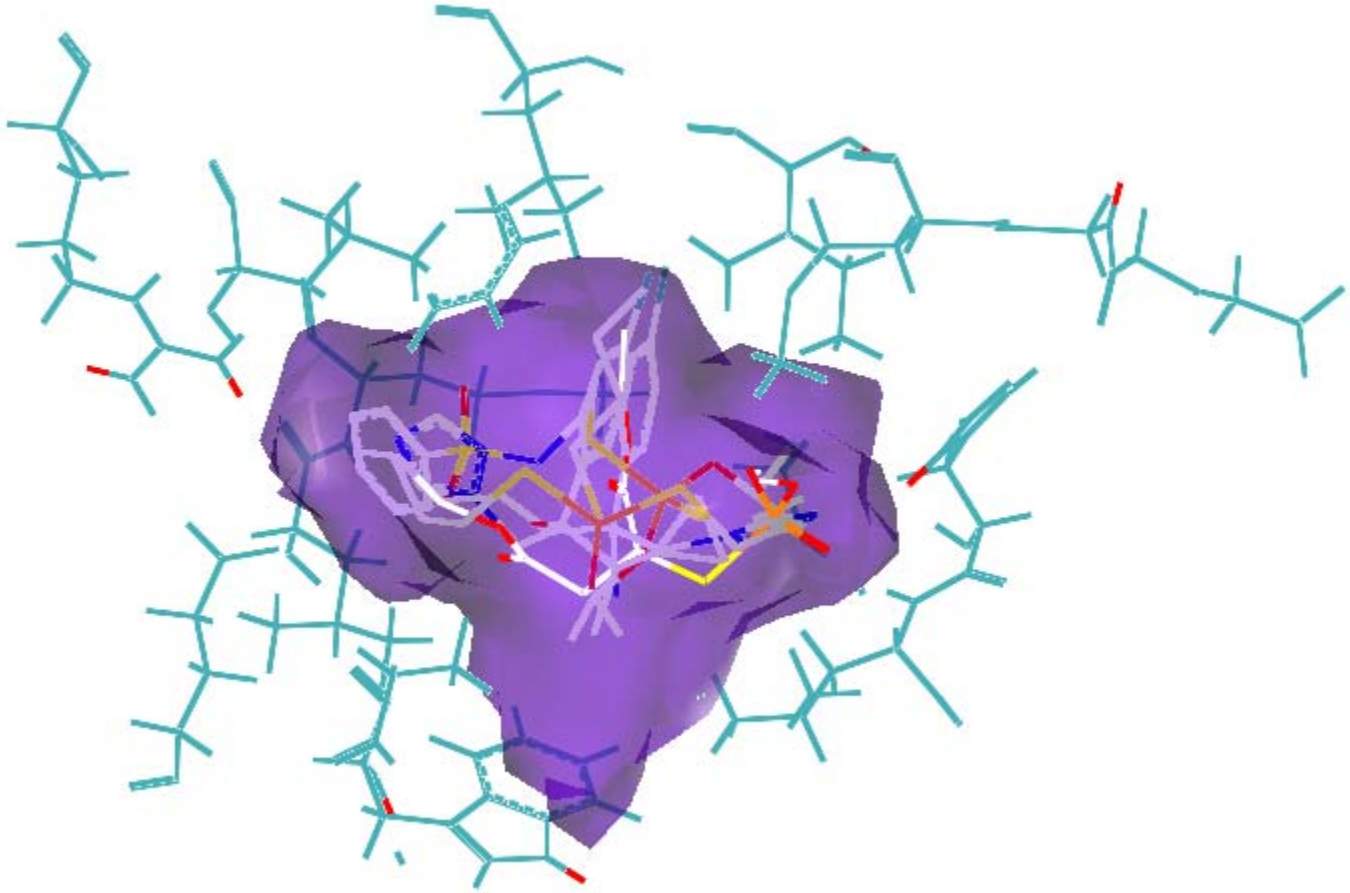
*** Smaller value denotes greater shape similarity to the query**

DES (known ERα agonist) was used as a query to search the Shape Signatures database of ToxCast chemicals.

Among the top 5 Toxcast chemicals ranked by ShapeSigs, 4 out of 5 molecules were reported as ERα *active* in the Attagene ATT_ERα_TRANS assay.

The top 5 hits, their experimental activities, ShapeSigs and docking scores are shown.

Docked Conformations of 4 Hits Predicted by Shape Signatures



overlapped docked conformations of ERα binding pocket with top 4 ShapeSigs hits, plus DES

Summary & Observations

- BSA is a promising approach for continued study of the Toxcast datasets
- BSA profiles show modest level of association (67%) with Chemical Descriptor profiles for the Attagene TA assay results
- Shape Signatures yielded encouraging results from screening for ERα agonists
- Classification models for two nuclear receptors (ERα and PXR) based on various Chemical Descriptors sets performed similarly well (Q_{10CV} (%) = 72-76%)
- Present study is preliminary

Points to Ponder

- Effect of assay conditions on test chemicals (chemical rxns? enzymatic rxns?)
- Effect of antagonists in Toxcast assays (TA assay gives signal only for agonists)
- Concentration corresponding to LEL value is different for each test chemical

Ongoing & Future Studies

- Apply SVM classifications to Toxcast results for other protein receptors
- Expand BSA profiles of data obtained from other ToxCast assays
- Explore other chemical descriptors sets & models
- Extend Shape Signatures modeling on ToxCast compounds

Acknowledgments

- **Funding**
  - USEPA funded environmental bioinformatics and Computational Toxicology Center (ebCTC) (STAR Grant RD-83272101)
- **Contributors**
  - Vladyslav Kholodovych, Ni Ai, Dmitriy Chekmarev, Panos Georgopoulos (UMDNJ)
  - Minjun Chen, Weida Tong (FDA-NCTR)

We invite you to visit our website

www.ebCTC.org for events, news, pubs, contacts

Thank You!

Viewpoints expressed here are the responsibility of the authors and do not necessarily reflect views of USEPA or its contractors.