Navigating the Chemistry/Biology Space:

A QSAR Adventure

Tudor I. Oprea, Marius Olah and Cristian Bologa

Office of Biocomputing

University of New Mexico School of Medicine

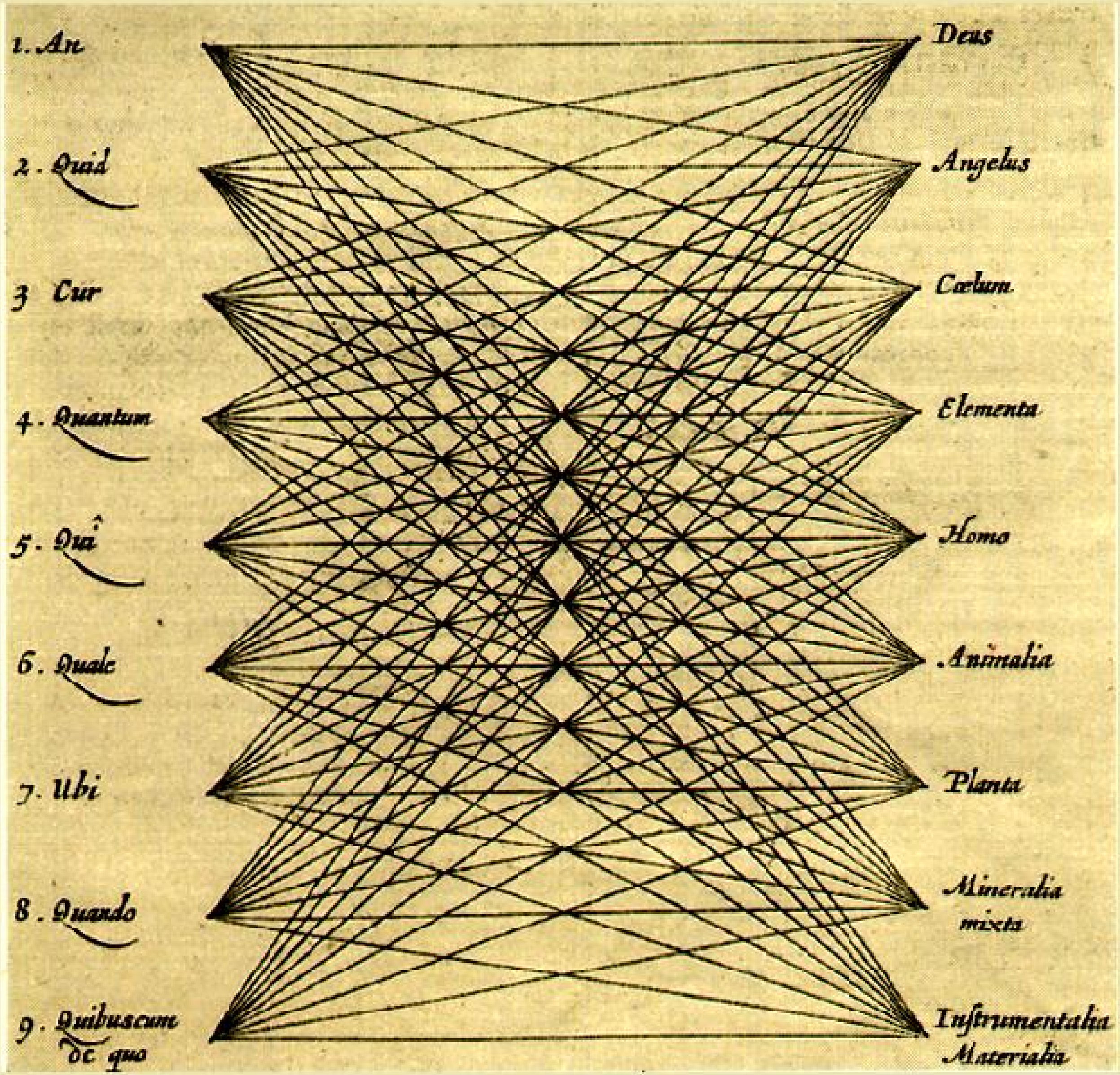
John Tallarico, Erik Brauner

Institute of Chemistry and Cell Biology

Harvard Medical School

Daylight MUG 04
Santa Fe, NM, 02/26/04

The University of New Mexico ♦ Health Sciences Center
SCHOOL OF MEDICINE



Athanasius Kircher:
Ars magna sciendi,
Amsterdam, 1669

Universal diagram for the formation of questions about every possible subject.

Universalschema zur Bildung von Fragen über alle möglichen Sachverhalte.

Schéma universel servant à poser des questions sur tous les sujets possibles.

Universellt diagram som beskriver konstruktion av frågor som beror alla möjliga ämnen.

Exploring Biological QSARs

- Started by Corwin Hansch in 1948
- Continued by Corwin Hansch to this day – by developing C-QSAR. Collaborative effort with Albert Leo, David Hoekman, Cynthia Selassie (Pomona College) and David Weininger (Daylight, Metaphorics, Green Chile Productions)
- Over 20,000 biological QSAR series have been entered in C-QSAR; based (mostly) on the Hansch equation – a monumental effort that started in 1962.
- The most amazing thing is that Corwin himself worked on these QSARs and, quite often, invented descriptors appropriate for the problem.
- It took him *at least* 48,000 hours to do this!!!
- C-QSAR is available from Biobyte and from Metaphorics

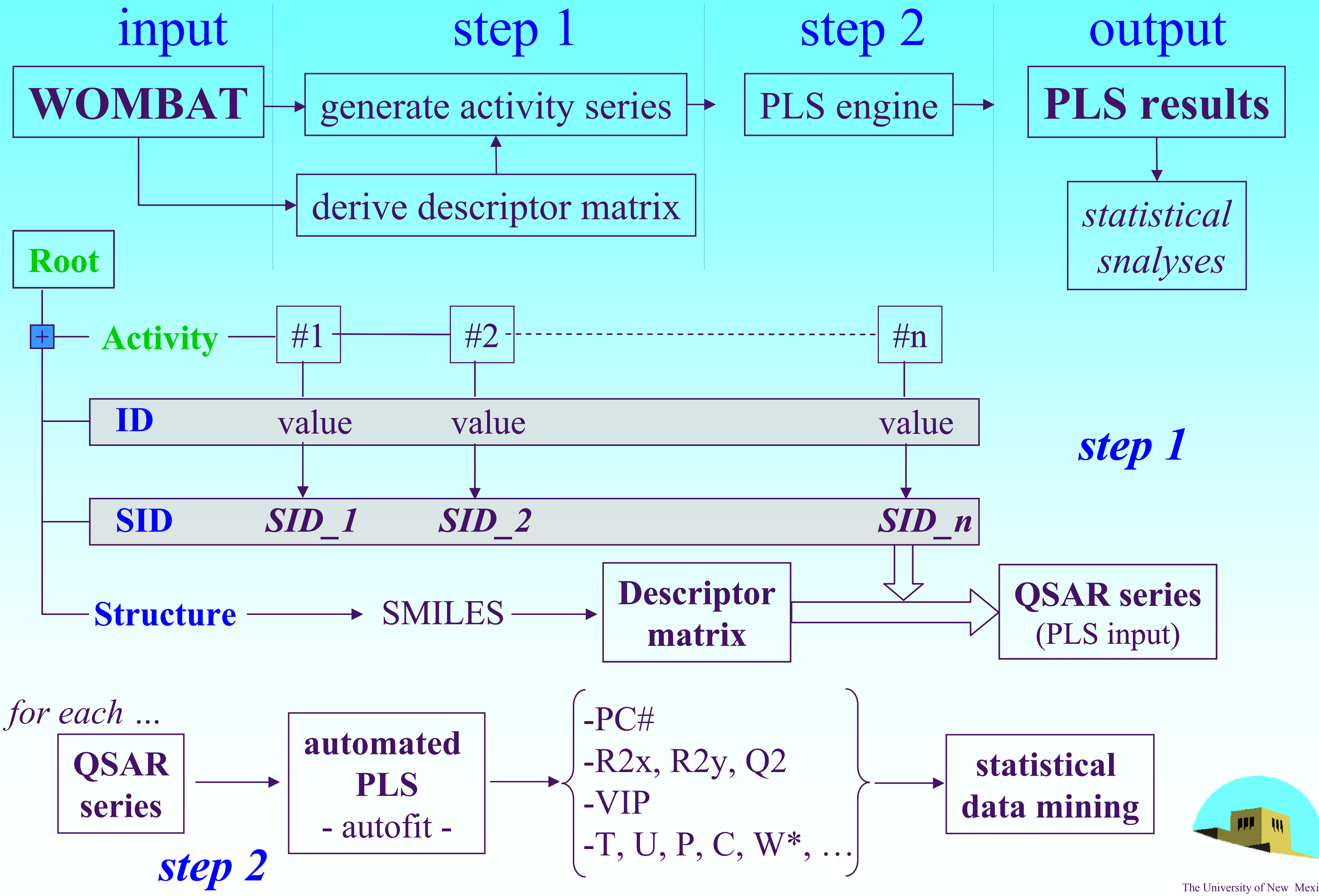
C-QSAR: An Inspiration

- C-QSAR is a unique asset in our field
- It applies a wide variety of descriptors related to π, σ, σ-M, σ-p, σ-I, σ*, Swain/Lupton, CLOGP, CMR, STERIMOL, etc.
- It offers a unified view over a vast bio-QSAR area
- It prompted the question: given a biological series, where do we begin to derive the QSAR?
- Can we start with any *a priori* assumptions about it?

C-QSAR: An Inspiration (2)

- Everyone has a pet descriptor set or a pet method
- Mine used to be SaSA (thanks to Vladimir Sherbukhin and Thomas Olsson, AZ) – a blend of 2D descriptors not unlike the 2D you get from MOE, QsarIs and other stooges
- We attempted to provide SaSA with a balanced view of the chemical universe
- We also wanted to have a more chemistry-oriented feel of the QSAR Universe. So we generated SMARTS (inspired from MDL’s 320 keys) to find what is relevant to biological activities
- 3D descriptors (e.g., pharmacophores) are under consdieration

Automated PLS Engine Flowchart



input

step 1

step 2

output

WOMBAT

generate activity series

PLS engine

PLS results

derive descriptor matrix

statistical snalyses

Root

Activity

#1

#2

#n

ID

value

value

value

step 1

SID

SID_1

SID_2

SID_n

Structure

SMILES

Descriptor matrix

QSAR series
(PLS input)

for each …

QSAR series

automated PLS
- autofit -

-PC#
-R2x, R2y, Q2
-VIP
-T, U, P, C, W*, …

statistical data mining

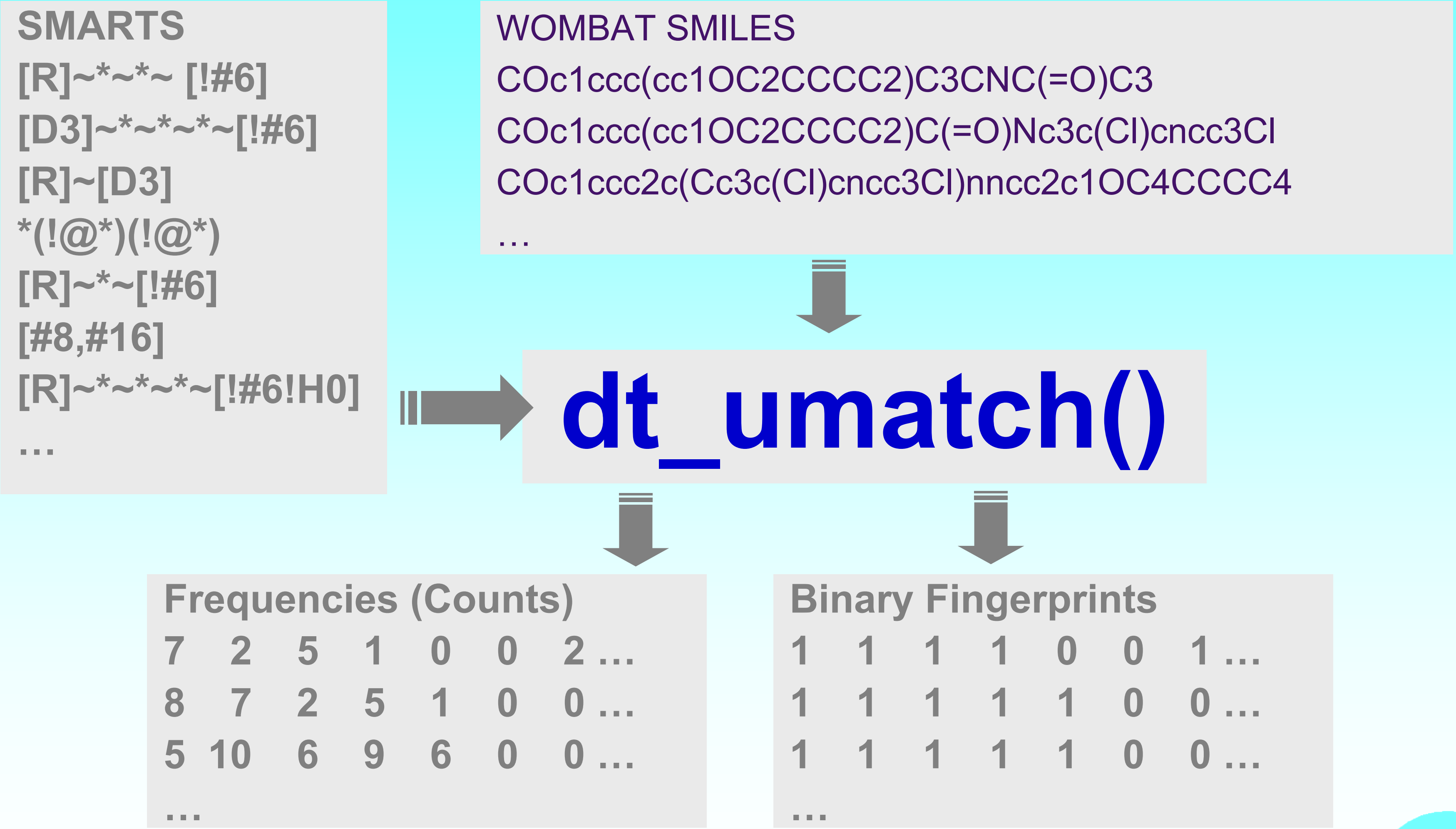
step 2

The University of New Mexico
SCHOOL OF MEDICINE

Initial Set of Descriptors

- Despite my kantian thirst for *a priori* reasoning (P.K. Dick calss them "precog"), I have to admit… …we had to start somewhere!
- As any psychologist can tell you: people often re-visit familiar places or situations (in memory or in 'reality')
- Hence, we decided to start with TCP, SaSA-like descriptors: the usual topological indices [Wiener, Randic, Motoc, Balaban, Kier Chi (p & c), Kier & Hall (3 of them)], atom counts [N, O, X, C, P_at, NP_at, etc.], hydrogen-bond counts [SMARTS definitions], Daylight's PCModels [CMR, CLogP], some electronic descriptors [Gasteiger charges plus Huckel MO info] and some 'complexity' [flexible bonds, rings, etc.]
- We added the 320 MDL keys produced by John and Norah MacCuish at Mesa Analytics and Computing LLC [OEChem based – data not shown]
- We added SMARTS inspired from MDL 320 keys and the WOMBAT patterns, produced by Vera Povolna and David Weininger at Metaphorics LLC

Fingerprints and Frequencies



SMARTS

[R]~*~*~ [!#6]

[D3]~*~*~*~[!#6]

[R]~[D3]

(!@)(!@*)

[R]~*~[!#6]

[#8,#16]

[R]~*~*~*~[!#6!H0]

…

WOMBAT SMILES

COc1ccc(cc1OC2CCCC2)C3CNC(=O)C3

COc1ccc(cc1OC2CCCC2)C(=O)Nc3c(Cl)cncc3Cl

COc1ccc2c(Cc3c(Cl)cncc3Cl)nncc2c1OC4CCCC4

…

dt_umatch()

Frequencies (Counts)

7 2 5 1 0 0 2 …

8 7 2 5 1 0 0 …

5 10 6 9 6 0 0 …

…

Binary Fingerprints

1 1 1 1 0 0 1 …

1 1 1 1 1 0 0 …

1 1 1 1 1 0 0 …

…

Code contributed by Tharun Kumar Allu (UNM)

The University of New Mexico
SCHOOL OF MEDICINE

WOMBAT Patterns

- Dave Weininger wrote a SMARTS generator starting from a SMILES that was hand-picked by Vera Povolna to match a *specific* (not the maximum common) substructure for each WOMBAT series
- These SMARTS are intended to capture the unique biological profile for each series – on occasion 2 such SMARTS were defined; note that hydrogens are matched exactly as defined in the series

[CH3]-[OH0]-[cH0]:1:[cH1,cH0]:[cH0]:2-[CH2]-[NH0](-[NH0]=[CH0](-[cH0]:2:[cH1]:[cH1]:1)-[CH2]-[cH0]:3:[cH0](:[cH1]:[nH0]:[cH1]:[cH0]:3-[ClH0])-[ClH0])-[CH0,SH0,CH1]=[OH0]

[CH2]-[CH2]-[NH0](-[CH2]-[CH2])-[CH2]-[CH2]-[OH0,SH0]-[cH0]:1:[cH1]:[cH1]:[cH0](:[cH1] :[cH1]:1)-[CH1]-2-[CH1](-[CH0,CH2]-[OH0]-[cH0]:3:[cH1]:[cH0](:[cH1]:[cH1]:[cH0]-2:3)-[OH0,OH1])-[cH0]:4:[cH1]:[cH1]:[cH1]:[cH1]:[cH1]:4

[OH1]-[CH0](=[OH0])-[CH2]-[CH1,CH2]-[NH1]-[CH0](=[OH0])-[CH2]-[NH1,NH0]-[CH0] (=[OH0])-[CH2,CH1,NH0]-[CH2]-[CH2]-[cH0]:1:[nH0]:[cH0]:2-[NH1]-[CH2]-[CH2]-[CH2]-[cH0]:2:[cH1]:[cH1]:1

[OH1]-[CH0](=[OH0])-[CH2]-[CH1,CH2]-[NH1]-[CH0](=[OH0])-[CH2]-[NH0]-1-[CH0](-[CH1](-[CH2]-[CH2]-1)-[CH2]-[CH2]-[cH0]:2:[nH0]:[cH0]:3-[NH1]-[CH2]-[CH2]-[CH2]-[cH0]:3:[cH1]:[cH1]:2)=[OH0]

[NH2]-[CH2]-[CH2]-[CH2]-[NH1]-[CH2]-[CH2]-[CH2]-[CH2]-[NH1]-[CH2]-[CH2]-[CH2]-[NH1]-[CH0,SH0]=[OH0]

- Provides interesting associations in FEDORA
- Inspired us to generate our own set of fingerprints

SUNSET MOLECULAR DISCOVERY LLC

Qriteria for QSAR Adventures

- Each series had minimum 25 compounds
- Each activity within a series was treated separately (even if Ki and IC50 values were provided for the same target)
- Each set of descriptors (e.g., TCP, FP500s, FQ500 and MDL 320) were computed separately for all series.
- Combinations of the above were produced, care being exercised about scaling (FPs were not centered to UV) and block-scaling
- We looked at the q2, using 3 cross-validation methods: LOO, CV7 and CV2; we considered q2>=0.3 worth looking at.
- We traced variables by looking at their VIP under CV7 and CV2
- We were interested what series ‘behave’ well, what series do not
- At this point we still do not examine individual series manually
- One immediate lesson: leave out leave-one-out (quoting Bob Sheridan from Merck)

QSAR Statistics

Stats	2D CV7	2D CV2	F500, CV7	F500, CV2	Q500, CV7	Q500, CV2
# QSARS	255	74	422	204	546	285
Md, # Cpds	38.5	48	40	47	38	42
Md, # Desc.	80	80	168	170	238	243
Md, #PCs	2	2	3	3	2	2
Md, R2(X)	0.6015	0.599	0.81	0.795	0.374	0.3325
Md, R2(Y)	0.681	0.685	0.762	0.764	0.793	0.783
Md, Q2(Y)	0.466	0.438	0.493	0.48	0.487	0.452

- Out of 1633 QSARs, only a fraction show significant Q2 (above 0.3) with the given descriptor sets – as noted in the #QSARs column.
- R2(X) shows how well the descriptors explain the X-block in a multivariate sense
- R2(Y) and Q2(Y) are more traditional QSAR measures.
- Q500 (the SMARTS counts) outperform the other methods – this was intended, since some SMARTS are designed to capture pharmacophore information
- Q500 is a blend between 2D and 3D, better than F500 since it is quantitative).

Trivial (?) 2D-QSARs

Series	N	K_320	A_320	R2Y_320	Q2_320	K_F500	A_F500	R2Y_F500	Q2_F500
SID_260	30	50	3	0.875	0.713	20	3	0.874	0.717
SID_460	38	135	3	0.895	0.687	44	3	0.848	0.664
SID_1563_2	109	148	5	0.911	0.784	50	4	0.688	0.517
SID_1627	42	72	2	0.717	0.568	23	3	0.799	0.546
SID_1640	114	238	4	0.874	0.787	79	4	0.836	0.738
Series	**N**	**K_Q500**	**A_Q500**	**R2Y_Q500**	**Q2_Q500**	**K_TCP**	**A_TCP**	**R2Y_TCP**	**Q2_TCP**
SID_260	30	56	1	0.817	0.681	77	1	0.786	0.656
SID_460	38	82	2	0.83	0.639	82	3	0.832	0.508
SID_1563_2	109	81	4	0.839	0.718	84	4	0.809	0.65
SID_1627	42	69	1	0.745	0.562	80	1	0.68	0.603
SID_1640	114	89	2	0.893	0.779	85	2	0.838	0.801

SID_260: A.S. Tasker et al., J. Med. Chem. 40, 1997, 322-330 – endothelin antagonists
SID_460: T. Su, et al., J. Med. Chem. 40, 1997, 4308-4318 – fibrinogen (GP IIb/IIIa) antagonists
SID_1563_2: A. Scozzafava et al., J. Med. Chem. 43, 2000, 292-300 – carbonic anhydrase II antagonists
SID_1627: B.C. Bookser, et al., J. Med. Chem., 43, 2000, 1495-1507 – AMP deaminase inhibitors
SID_1640: C.T. Supuran, et al. J. Med. Chem., 43, 2000, 1793-1806 – thrombin inhibitors

Series	N	K_320	A_320	R2Y_320	Q2_320	K_F500	A_F500	R2Y_F500	Q2_F500
SID_1530	29	65	3	0.968	0.834	22	6	0.948	0.736
Series	**N**	**K_Q500**	**A_Q500**	**R2Y_Q500**	**Q2_Q500**	**K_TCP**	**A_TCP**	**R2Y_TCP**	**Q2_TCP**
SID_1530	29	38	5	0.989	0.87	71	1	0.442	0.272

SID_1530: L. Amat, et al., Med. Chem., 42, 1999, 5169-5180 – trypsin inhibitors (quantum similarity)

Why 3D QSAR is Needed

Series	**N**	**K_320**	**A_320**	**R2Y_320**	**Q2_320**	**K_F500**	**A_F500**	**R2Y_F500**	**Q2_F500**
SID_284	48	137	0	0	0	44	0	0	0
SID_287	30	65	0	0	0	22	0	0	0
SID_317	50	162	3	0.796	0.527	56	0	0	0
SID_1056	49	188	0	0	0	49	0	0	0
Series	**N**	**K_Q500**	**A_Q500**	**R2Y_Q500**	**Q2_Q500**	**K_TCP**	**A_TCP**	**R2Y_TCP**	**Q2_TCP**
SID_284	48	65	0	0	0	80	1	0.371	0.19
SID_287	30	60	1	0.512	0.127	80	0	0	0
SID_317	50	85	2	0.7	0.169	83	0	0	0
SID_1056	49	76	1	0.599	0.351	83	0	0	0

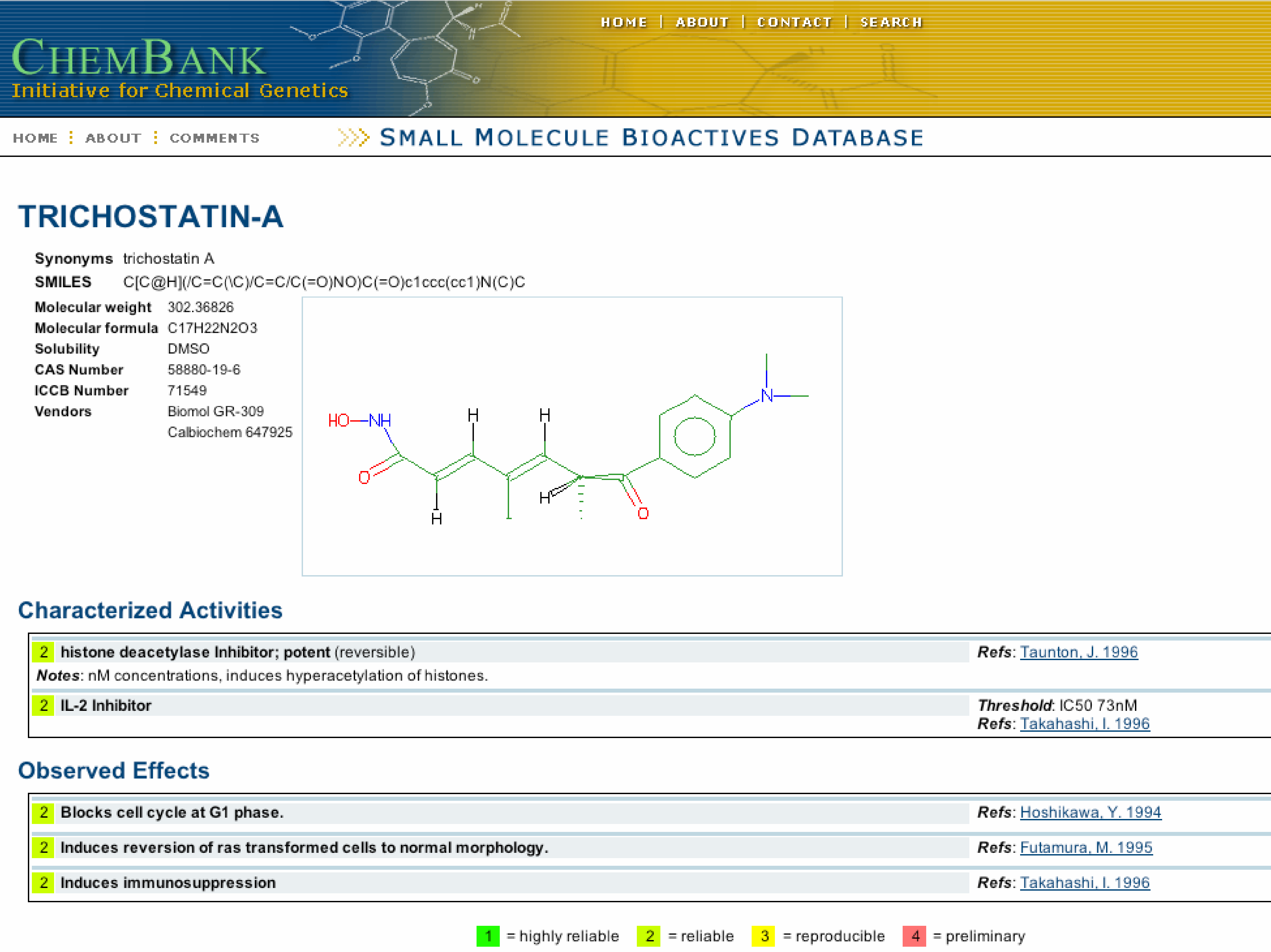
SID_284:	S. Sicsic et al., J. Med. Chem. 40, 1997, 739-748 – Melatonin (GPCR) antagonists
SID_287:	J. Nilsson et al., J.Med.Chem., 40, 1997, 833-840 – Dopamine D3 receptor antagonists
SID_317:	M. Pastor et al., J. Med. Chem. 40, 1997, 1455-1464 – Glycogen phosphorylase b inhibitors
SID_1056:	M. K. Holloway et al., J.Med.Chem. 38, 1995, 305-317 – HIV protease inhibitors

VIP Criteria (2D and SMARTS)

2D CV7	2D CV2	F500, CV7	F500, CV2	Q500, CV7	Q500, CV2
CMR	NrBonds	[CH3]-*~*~[R]	*-!:[a]:*:*:[a]-!:*	[D3]~[R]	[D3]~[R]
Polarizability	NrAtoms	[CH3]-*~[R]	*!:[a]:*:*:[a]!:*	[!$(*#*)&!D1]-!@[!$(*#	[!$(*#*)&!D1]-!@[!$(*#*)
MolVol2D	CMR	[CH3]-*~*~[a]	[CH3]-*~*~[R]	[!#6]~*~*~[R]	[D3]~*~*~*~[!#6]
Randic_index	Inform_content	[$([#6]-!@[A!C!H]-!@[#	[$([#6]-!@[A!C!H]-!@[#	[D3]~*~*~*~[!#6]	[!#6]~*~[R]
NrBonds	Polarizability	[#9,#17,#35,#53]	[CH3]-*~[R]	[!#6]~*~[R]	[R]
NrAtoms	Kier_Chi0	[#9,#17,#35,#53]~*(~*)	[CH3]-*~*~[a]	[R]	[#8,#16]
Carbon_count	Sum(N,O,P,S)	[CH3]-*~[a]	[$([#8X2v2]([#6])[#6])]	*-*-[R]~*:[a]	[!#6]~*~*~[R]
Kier_Chi0	Randic_index	*-!:[a]:*:*:[a]-!:*	[#9,#17,#35,#53]	[#8,#16]	[!a](=*)~*~*~*~[R]
Kier_Chi5p	K&H_Kappa1	[#6!H0]~@[#6!H0]~@[#	[#9,#17,#35,#53]~*(~*	*(!@*)(!@*)	*-*-[R]:[a]
Kier_Chi3p	MolVol2D	[!#6]~[CH3]	[!#6]~[CH3]	[R](-*(-*))~*~*~[a]	[R](-*(-*))~*~*~*~[a]
MW	Motoc_index	*!:[a]:*:*:[a]!:*	[CH3]-*~[a]	*-*-[R]:[a]	[R](-*(-*))~*~*~[a]
NonPolatoms	Wiener_index	*-!:[a]:*:[a]-!:*	[#7]~*~*~*~*~*~*~[#8]	[R](-*(-*))~*~[a]	*-*-[R]~*:[a]
TSA	TSA	[$([#8X2v2]([#6])[#6])]	*!@[#8]!@*	[R](-*(-*))~*~*~*~[a]	[R](-*(-*))~*~[a]
K&H_Kappa1	Kier_Chi2	[#7]~*~*~*~*~*~*~*~[	[#6!H0]~@[#6!H0]~@[	[#7]~*(~*)~*	[#6X3v4+0,#6X3v3+1,#
Kier_Chi2	Carbon_count	[#6!H0]~@[#6!H0]~@[#	*-!:[a]:*:[a]-!:*	[!a](=*)~*~*~*~[R]	*(!@*)(!@*)
Wiener_index	Kier_Chi3p	[R](-*(-*))~*~*~*~[R](-*(	[!a]=*~*~[D3]	*-*-[R]~*~*:[a]	*-*-[R]~*~*:[a]
Inform_content	Kier_Chi5p	*!@[#8]!@*	[#8]~[#6](~[#6])~[#6]	[#6X3v4+0,#6X3v3+1	[#6]~!@[#8]
Motoc_index	Graph_diameter	[#6!H0]~@[#6!H0]~@[#	[#9,#17,#35,#53]-[a]	[#6]~!@[#8]	[#7]~*(~*)~*
Kier_Chi6p	HMO_pi-energy	*~!@[CH2&!R]~!@*	[#9,#17,#35,#53]!@*@	[CX4v4]	[!a]=*~*~*~[D3]
NPSA	PSA	[#9,#17,#35,#53]!@*@	[#9,#17,#35,#53]-[R]	[D3]~[#7]	[D3]~[#7]

Summary of the QSAR Adventures

- LOO measures redundancy (data not shown); CV2 is too severe – thus limited small groups cross-validation (CV7) is better for model consistency
- Among the 2D descriptors (useful QSARs for 15.6% of the series), topological indices get 10 out of top 20 VIP counts.
- The MDL public set (320 FPs) appear, indeed, to be ‘drug-like’ (useful QSARs for 21% of the series).
- The F500 (25.8%) and Q500 (33.4%) appear to capture more QSARs compared to the other sets. Q500 is blending quantitative (2D-like) and qualitative (fingerprint-like) descriptors, hence it is more successful
- 3D descriptors are likely to provide additional, useful QSAR models



CHEMBANK

Initiative for Chemical Genetics

SMALL MOLECULE BIOACTIVES DATABASE

TRICHOSTATIN-A

Synonyms	trichostatin A
SMILES	C[C@H](/C=C(\C)/C=C/C(=O)NO)C(=O)c1ccc(cc1)N(C)C
Molecular weight	302.36826
Molecular formula	C17H22N2O3
Solubility	DMSO
CAS Number	58880-19-6
ICCB Number	71549
Vendors	Biomol GR-309 Calbiochem 647925

Characterized Activities

2	**histone deacetylase Inhibitor; potent** (reversible) ***Notes***: nM concentrations, induces hyperacetylation of histones.	***Refs***: Taunton, J. 1996
2	**IL-2 Inhibitor**	***Threshold***: IC50 73nM ***Refs***: Takahashi, I. 1996

Observed Effects

2	**Blocks cell cycle at G1 phase.**	***Refs***: Hoshikawa, Y. 1994
2	**Induces reversion of ras transformed cells to normal morphology.**	***Refs***: Futamura, M. 1995
2	**Induces immunosuppression**	***Refs***: Takahashi, I. 1996

🟩 1 = highly reliable 🟨 2 = reliable 🟨 3 = reproducible 🟥 4 = preliminary

Top 200 Drugs Database

- **200 drugs**
  - 3 drugs with 3 active ingredients (Act.Ing); 34 with 2 Act.Ing; 163 ‘singles’
  - 73 drugs include non-chiral Act.Ing, 97 drugs include chiral Act.Ing, 46 drugs include racemic Act.Ing
  - Manufacturers Ranking: Pfizer (16); GSK (15); AstraZeneca (7); BMS (7); Merck (6); Lilly (5); Ortho-McNeill (5); Schering (5); Abbott (5); Aventis (4); Novartis (4); Wyeth (4); 76 are ‘various’
- **160 unique structures**
  - Act.Ing Ranking: Ethinyl-Estradiol (9); Acetaminophen (6); Amoxicillin (6); HCTZ (6); Loratadine (4); Metformin (4); Lisinopril (4)
  - Chemical Class Ranking: Steroids (18); Phenyl-ethyl-amine (14); AINS (7); BZP (7); ‘Pril’ (7); ‘Tricyclics’ (6); beta-lactam (5); Opiate (5); ‘prazole’ (5)
  - Therapeutic Category Ranking: Antihypertensive (25); Antibacterial (18); Antidepressant (10); Antiinflammatory (10); Analgesic (9); Antianginal (8); ‘estrogen’ (8); Antiarrhythmic (7); Antiulcerative (7); Contraceptive (6);

As published at http://www.rxlist.com/top200.htm

SUNSET MOLECULAR DISCOVERY LLC

Top 158 Drugs 1D Projection Method

- **158 (unique) drugs**
  - From the 160 unique structures, we excluded KCl and Insulin
  - Esomeprazole is confounded with Omeprazole due to improper chiral perception of the R-S(=O)-R1 function in Daylight SMILES [this is fixed in OpenEye’s OEChem and in the coming SMILES 5.0 from Daylight]
  - MDL 320 keys were generated using MESA Analytics Software
- **PCA on MDL 320 keys**
  - PCA (no scaling and centering) was performed on the 158x320 input matrix; ca 20% of the keys were excluded due to zero variance; 5 PCs were extracted after cross-validation in SIMCA
- **Tversky Similarity indices**
  - The full similarity matrix based on Tversky asymmetric similarity indices, where A->B differs from B->A
  - Tversky(A,B) = c / [(alpha) * a + (beta) * b + c] (asymmetric)
  - Where alpha = 1 – beta (typically alpha is 0.9 or 0.95)
    a : Unique bits turned on in molecule "A"
    b: Unique bits turned on in molecule "B"
    c: Common bits turned on in both molecule "A" and molecule "B"
    High Tversky (A,B) values imply that A "fits into" B

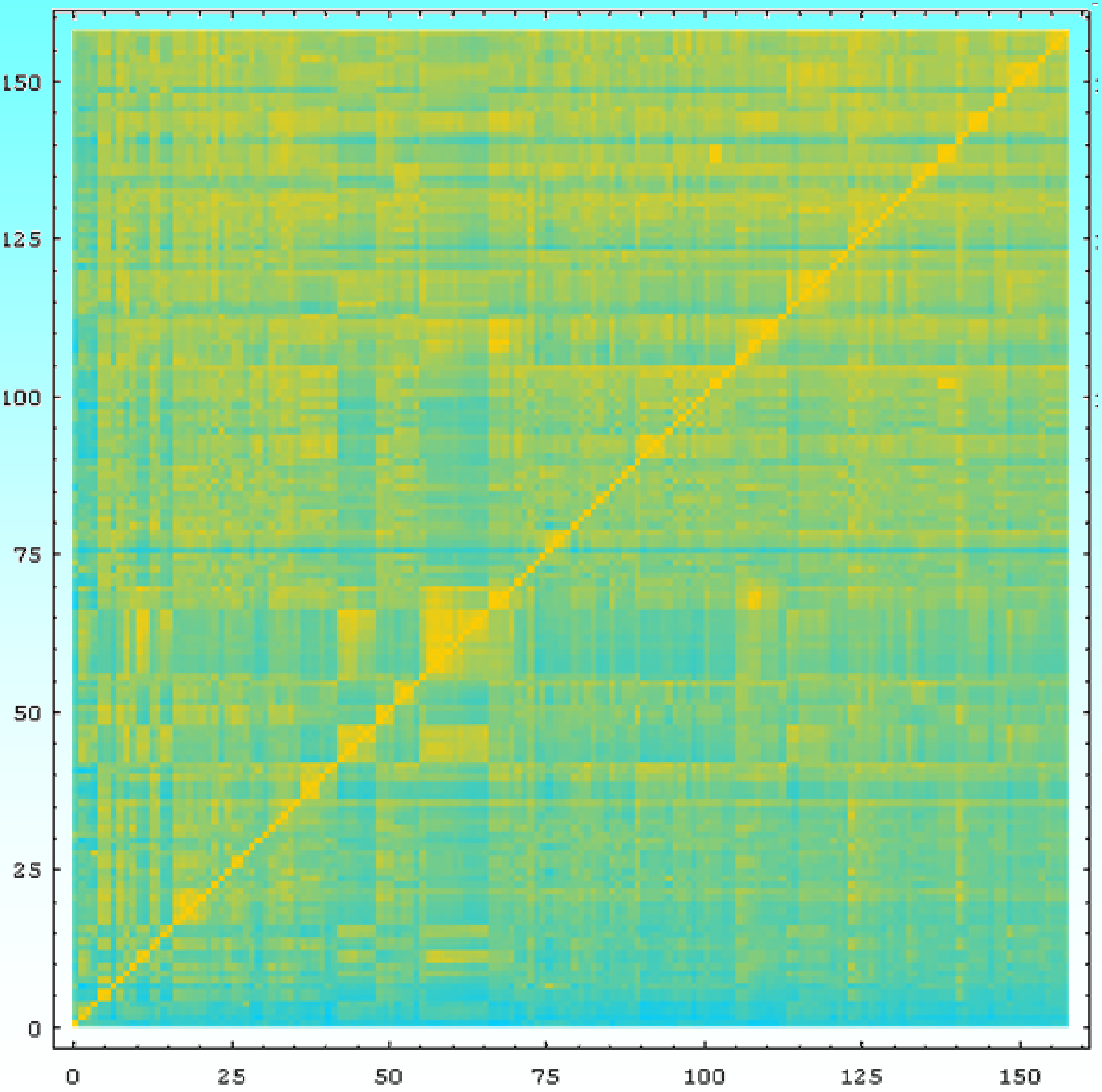
Tversky Similarity and Clustering code contributed by Norah and John MacCuish (MESA)

Top 158 Drugs 1D Projection Method

- **Clustering based on Tversky Similarity indices**
  - The full similarity matrix based on Tversky indices was used to cluster compounds at 0.85 cut-off;
  - Clustering (Asymmetric Taylor, non-disjoint) produced 23 clusters and 42 true singletons.
- **Ordering of the compounds in 1D**
  - The 23 cluster centroids and the 42 singletons were ordered according to their t1 (PCA) values; within each cluster, compounds were ordered according to their t1 values
- **Advantages of using 1D Tversky similarity on MDL320**
  - Fixed fingerprints (as opposed to the Daylight or Barnard ones) allow consistent mapping throughout all chemical space [this is also a disadvantage]
  - Structure similarity shows on the horizontal axis
  - Substructure similarity shows on the vertical axis
  - This is the prototype of the Similarity Navigator (SimNav)

Tversky Similarity and Clustering code contributed by Norah and John MacCuish (MESA)

Similarity Matrix on 158 Drugs

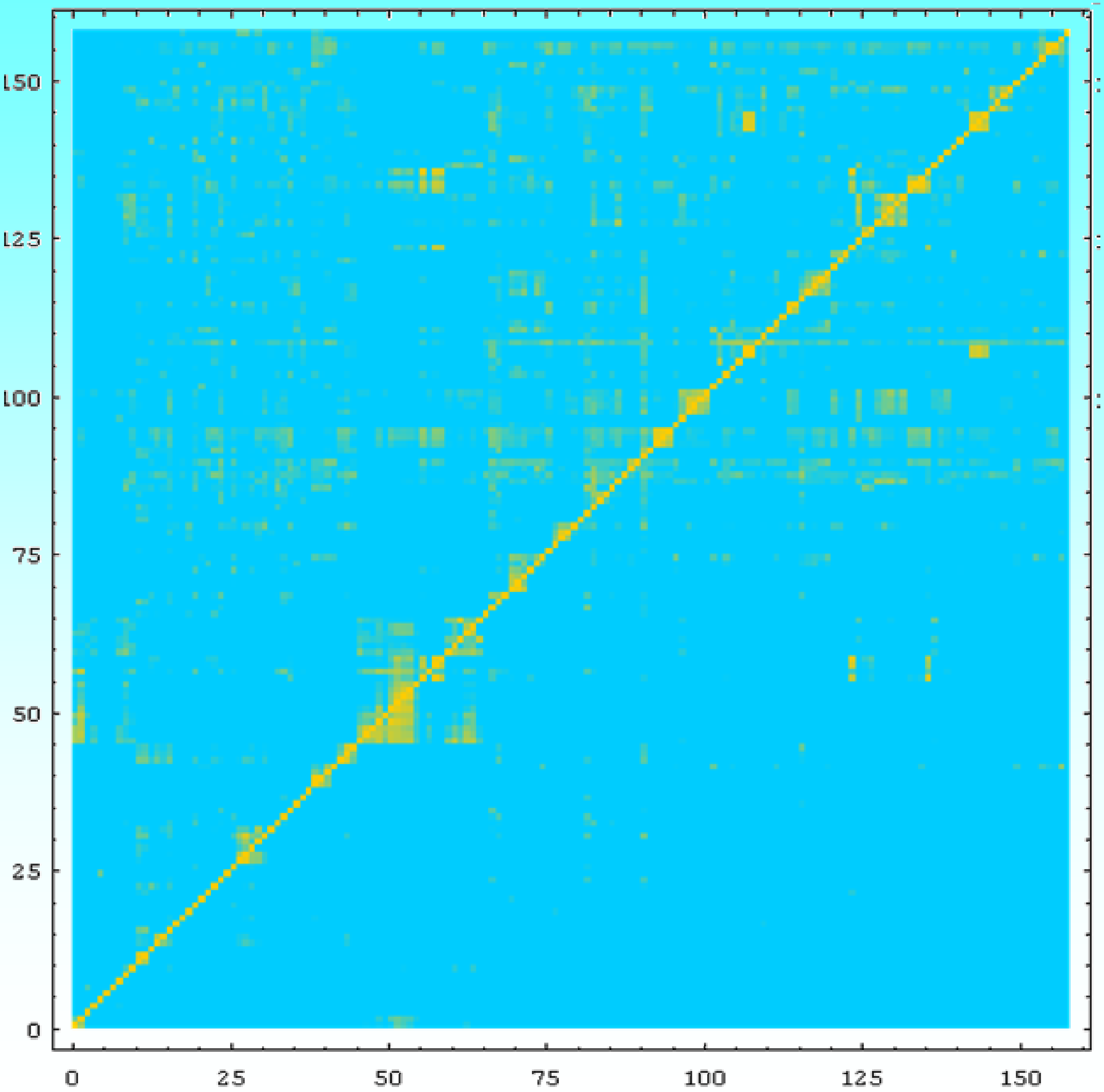


t1-t1
0-1 range

Tversky similarity on MDL 320 keys

The University of New Mexico
SCHOOL OF MEDICINE

Similarity Matrix on 158 Drugs

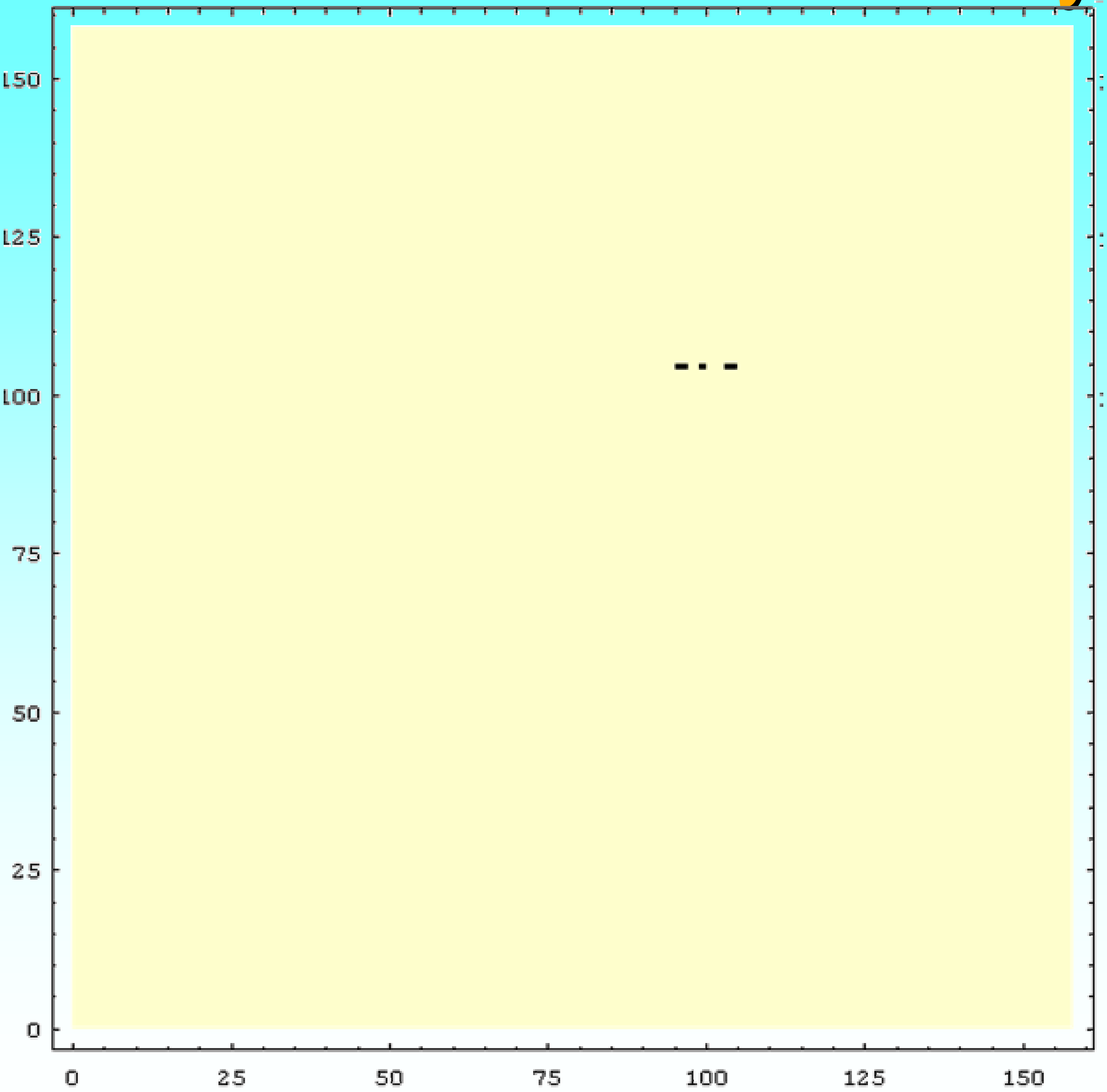


t1-t1
0.7-1 range

Tversky
similarity
on MDL
320 keys

The University of New Mexico
SCHOOL OF MEDICINE

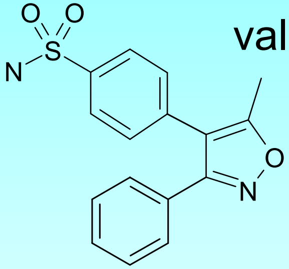
Sildenafil on the Similarity Navigator



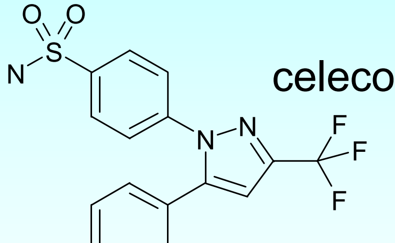
t1-t1

0.93-1 range

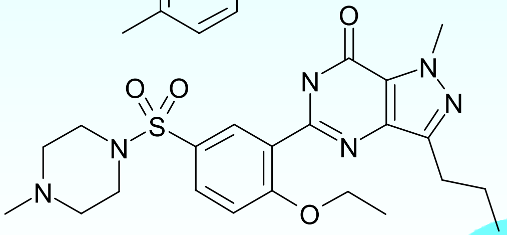
Tversky similarity on MDL 320 keys



valdecoxib

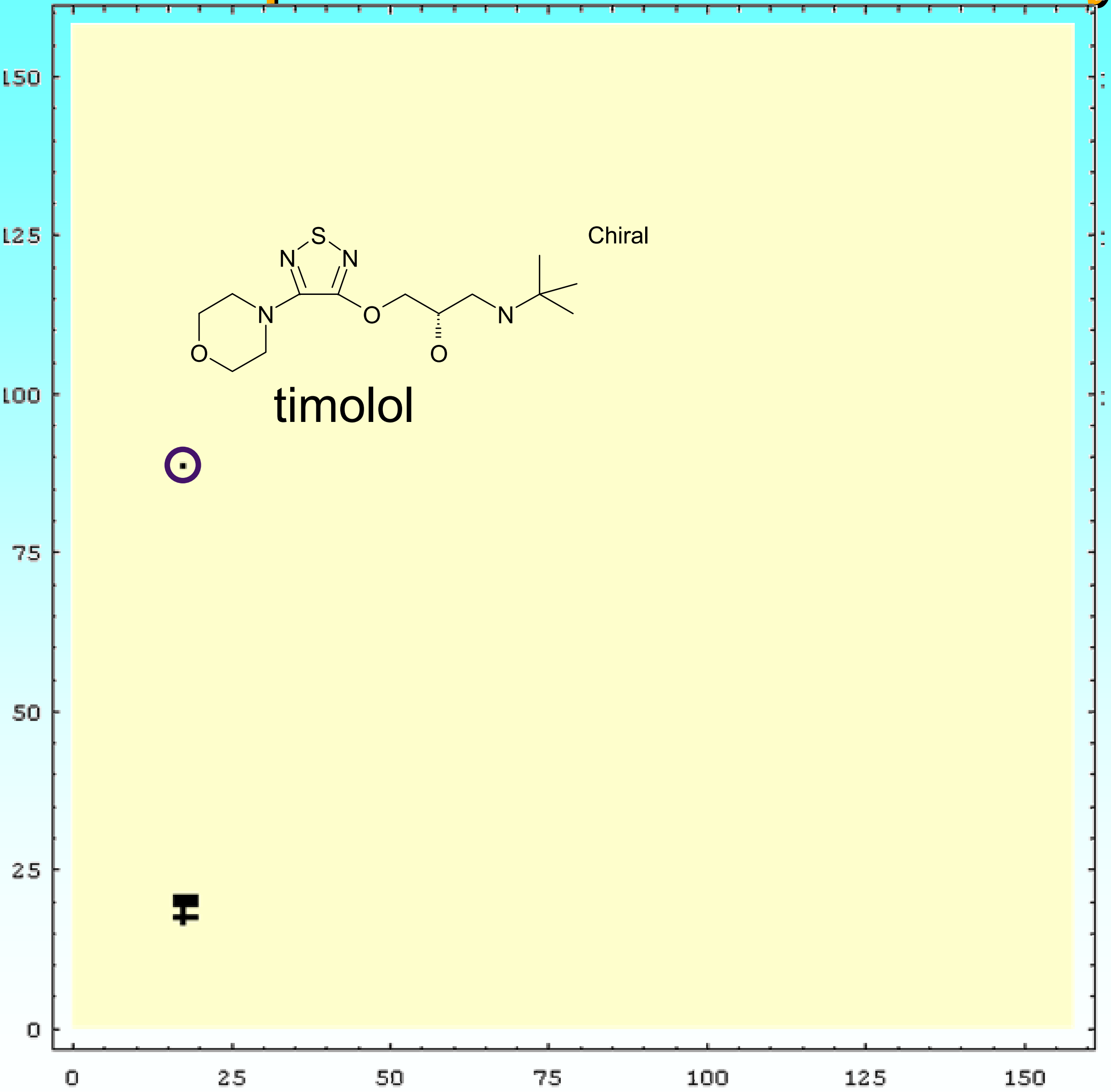


celecoxib



sildenafil

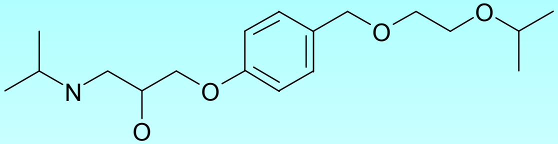
Metoprolol on the Similarity Navigator



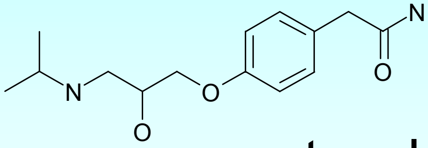
t1-t1

0.9-1 range

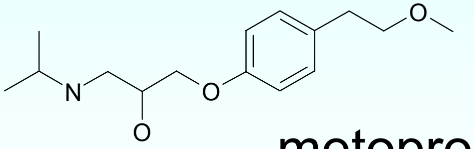
Tversky similarity on MDL 320 keys



bisoprolol



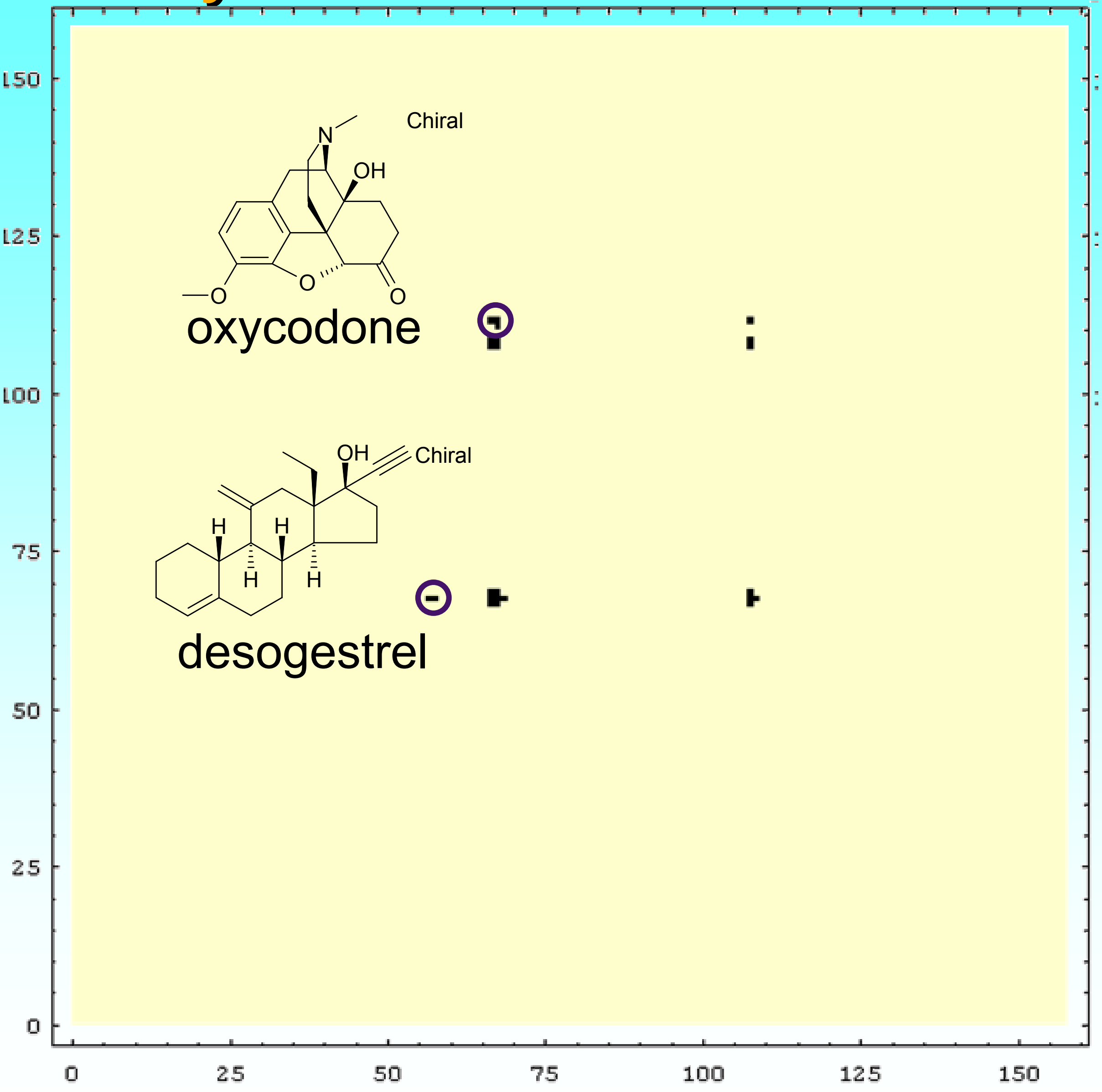
atenolol



metoprolol

The University of New Mexico
SCHOOL OF MEDICINE

Ethinyl-Estradiol on the Similarity Navigator

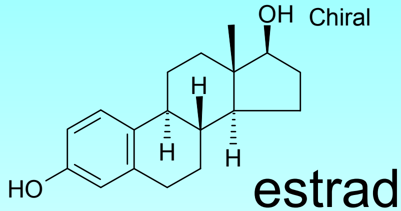


t1-t1

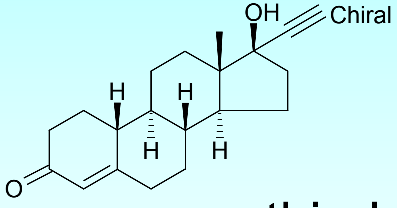
0.9-1 range

Tversky similarity

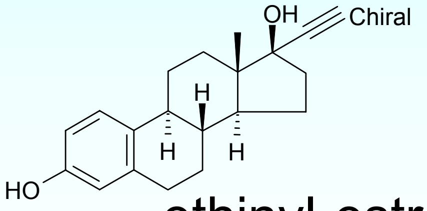
on MDL 320 keys



estradiol



norethindrone



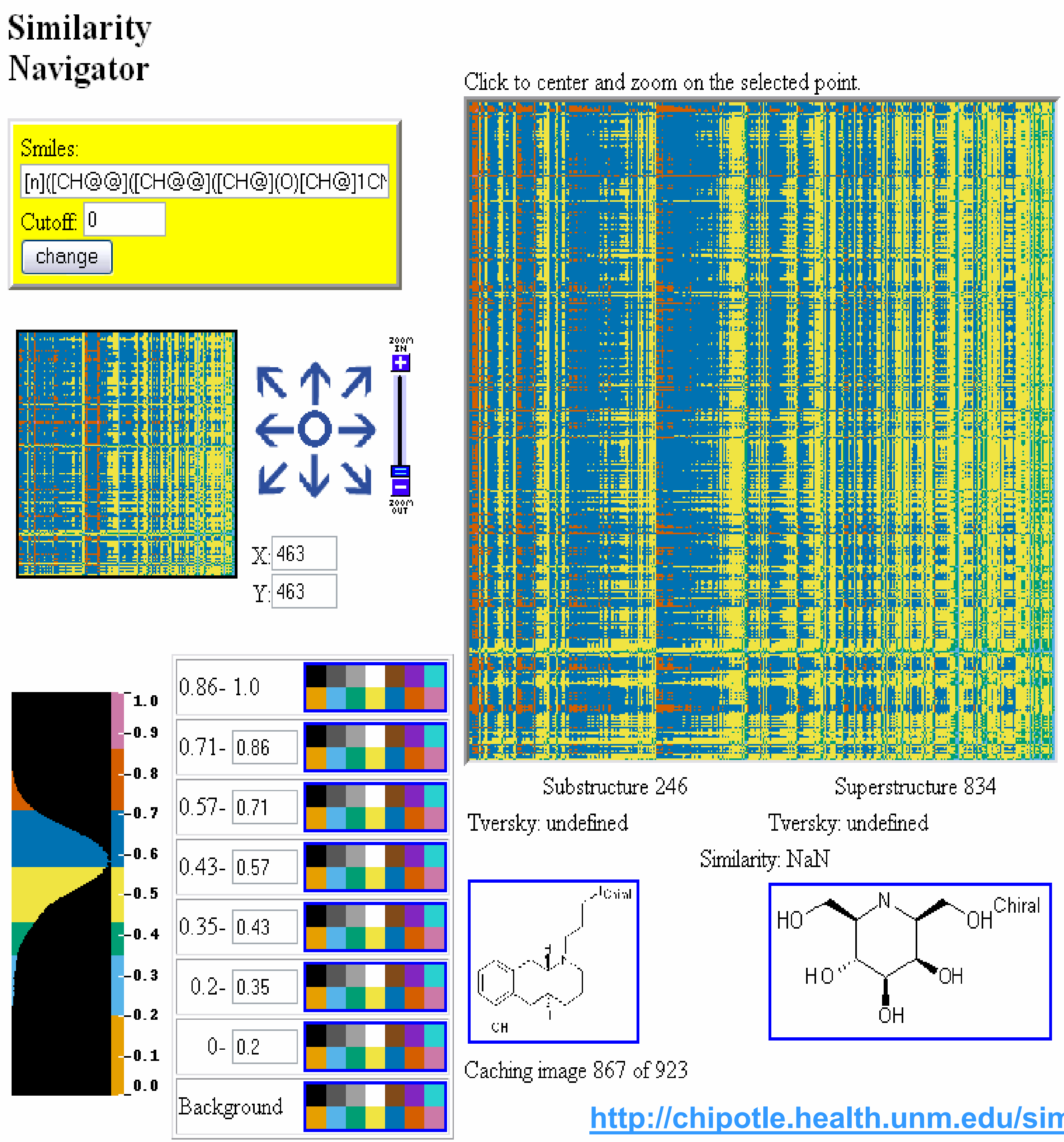
ethinyl-estradiol

SimNav: Extensions with WOMBAT

- **WOMBAT (WOrld of Molecular BioAcTivity)**
  - Over 76000 entries (>3000 papers, ~140000 activities) from literature
  - Developed by Sunset Molecular Discovery, marketed by Daylight, and available at Harvard's ICCB
  - Automated-QSAR ready
- **Extension for SimNav:**
  - Added the highest WOMBAT active for each target to the SimNav 1D-projection system
  - This represents 769 unique structures active on 549 targets (most of them IC50s, Kis or EC50s)
  - This enhances the chemical and biological diversity of the system – no info about *selectivity*
  - On-going development – to be tested at ICCB

Activity	Percentage
A2	3.311
appIC50	0.042
appKb	0.126
D2	1.299
EC50	**11.442**
IC50	**45.222**
Kb	0.335
Kd	0.126
Ki	**38.013**
Kii	0.042
Kis	0.042

SimNav prototype coded by Andrew Dalke (Dalke Scientific)



Similarity Navigator

Smiles: [n]([CH@@]([CH@@]([CH@](O)[CH@]1CN

Cutoff: 0

change

X: 463

Y: 463

0.86- 1.0	
0.71- 0.86	
0.57- 0.71	
0.43- 0.57	
0.35- 0.43	
0.2- 0.35	
0- 0.2	
Background	

Click to center and zoom on the selected point.

Substructure 246

Tversky: undefined

Superstructure 834

Tversky: undefined

Similarity: NaN

Caching image 867 of 923

SimNav: Prototype

- So far, SMILES driven (not for the faint of heart)
- The color code is for the color-blind
- The distribution, the zoom-in and the navigation bar are helpful
- The 2D pictures could improve
- Work in progress

http://chipotle.health.unm.edu/simnav/simnav.cgi

The University of New Mexico
SCHOOL OF MEDICINE

Acknowledgments

- Tharun Kumar Allu contributed the SMARTS count code
- Norah and John MacCuish (Mesa Analytics and Computing) provided MDL fingerprints & clustering
- Vera Povolna and Dave Weininger (Metaphorics) mapped unique SMARTS to each WOMBAT series
- Andrew Dalke (Dalke Scientific) produced the first prototype of SimNav

EuroQSAR 2004

Chair: Prof. Dr. Esin AKI ŞENER

sener@pharmacy.ankara.edu.tr

Co-Chair: Prof. Dr. İsmail YALÇIN

yalcin@pharmacy.ankara.edu.tr

Address for Correspondence:

Armoria Congress

armoria@euro-qsar2004.org

15th European Symposium on
Quantitative Structure-Activity Relationships
Istanbul / Turkey 05-10 September 2004
www.euro-qsar2004.org