Molecular Simulations Inc.

A Subsidiary of Pharmacopeia, Inc.

Automatic Pharmacaphore Development using Structure Based Focussing

C. M. Venkatachalam, Paul Kirchhoff and Marvin Waldman

Molecular Simulations Inc

A Subsidiary of Pharmacopeia Inc.

Problem of Pharmacaphore Development in Structure Based Focussing

Known:

3D structure of Protein

active site

A library of Compounds

- **A small of set with known activities**
- **A large library**

Problem:

What subset of a library of compounds can satisfactorily bind to the active site?

What pharmacaphores will bring the highly active compounds and reject inactive compounds?

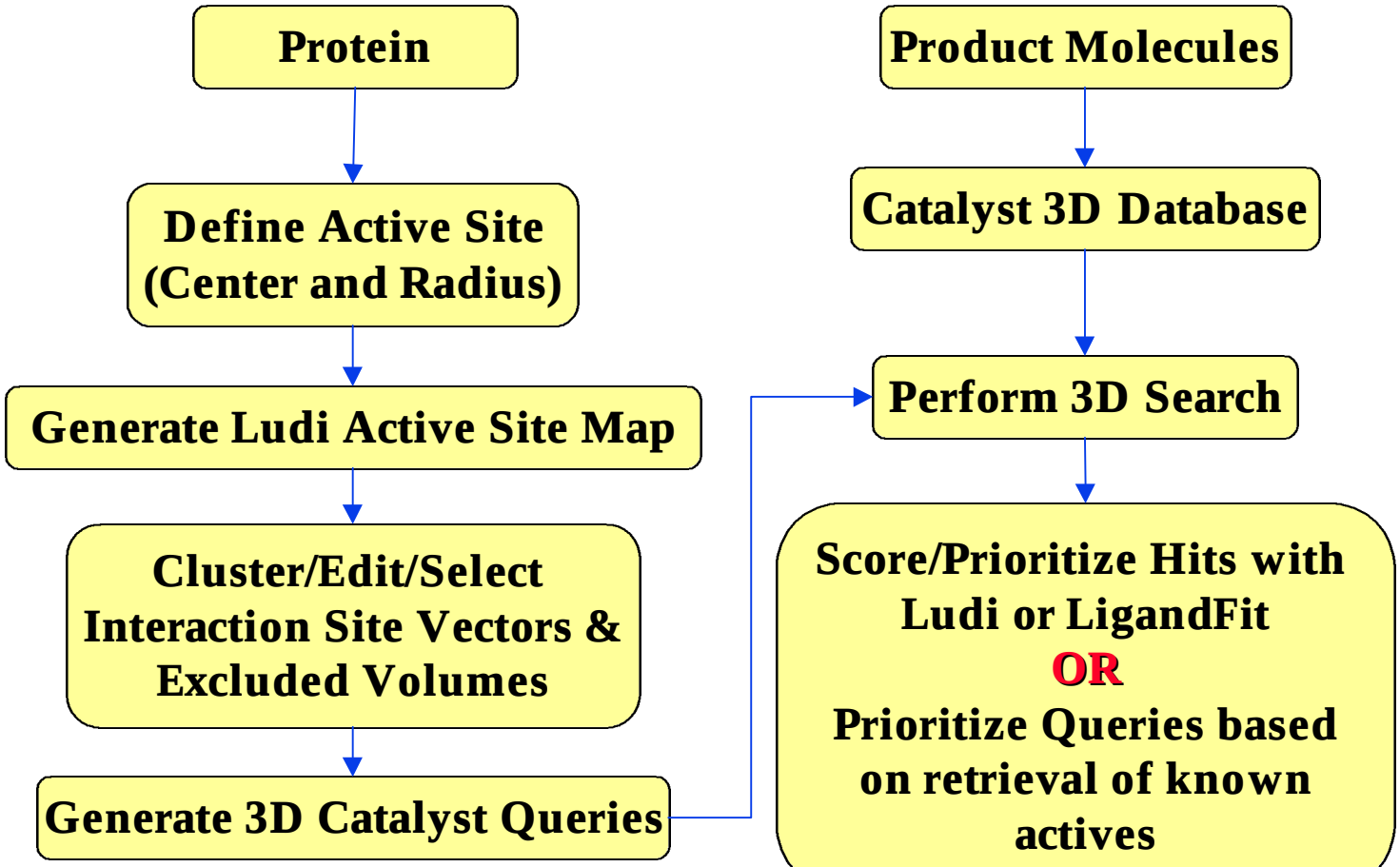
Two New Applications in Cerius2

Cerius2 Structure Based Focusing (C2*SBF)

Cerius2 LigandFit

A Subsidiary of Pharmacopeia Inc.

Cerius^2 Structure Based-Focusing



A Subsidiary of Pharmacopeia Inc.

Application of Cerius^2 Structure-Based Focusing to Estrogens

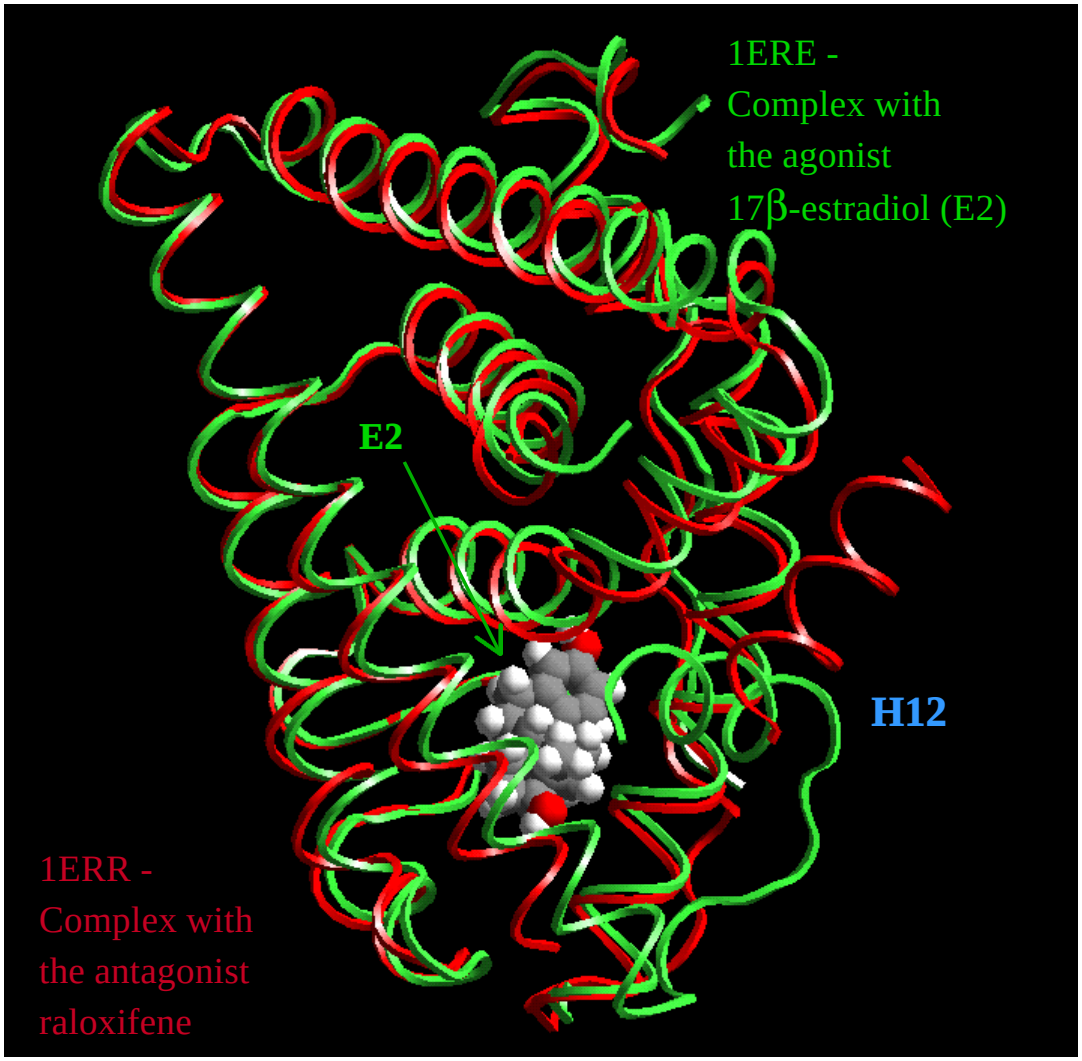
A project in collaboration with Affymax

Paul Kirchhoff

Estrogens are vital to the endocrine systems of humans and other vertebrates regulating the growth, development and homeostasis of a variety of tissues.

- Basic research
- As a target for the development of therapeutic agents
- Concerns that man-made chemicals may mimic endogenous hormones

Plasticity of the Estrogen Receptor



Brozozowski, A.M.; Pike, A.C.W.; Dauter, Z.; Hubbard, R.E.; Bonn, T.; Engström, O.; Öhman L.; Greene, G.L.; Gustafsson, J-A.; Carlquist, M. *Nature* **1997**, 389, 753-758.

Relative Binding Affinities (RBA)*

Ligand	RBA†	K_i (nM)‡	Ligand	RBA†	K_i (nM)‡
Diethylstilbestrol	468	0.04	2-hydroxy-E2	7	2.5
Hexestrol	302	0.06	Tamoxifen	7	3.4
Dienestrol	223	0.05	5-androstanediol	6	3.6
4-hydroxy-tamoxifen	178	0.10	Genistein	5	2.6
E2	100	0.13	3β-androstanediol	3	6
Coumestrol	94	0.14	Norethynodrel	0.7	14
ICI164384	85	0.2	4-androstenediol	0.5	23
Estrone	60	0.3	3α-androstanediol	0.07	260
17α-estradiol	58	0.2	Norethindrone	0.07	152
Nafoxidene	44	0.3	BisphenolA	0.05	195
Moxestrol	43	0.5	5α-dihydrotestosterone	0.05	221
Clomifene	25	0.9	Dehydroepiandosterone	0.04	245
β-zearalanol	16	0.8	Methoxyclor	0.01	1774
Estriol	14	1.4	Testosterone	<0.01	…
4-hydroxy-E2	13	1.0	Progesterone	<0.001	…

Steroids, synthetic estrogens, phytoestrogens, industrial chemicals, antiestrogens

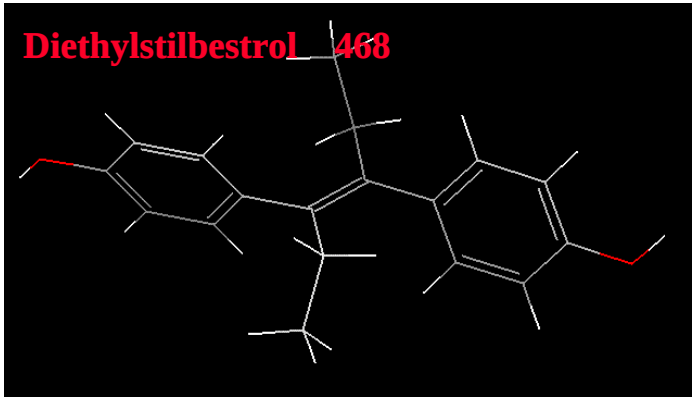
* Kuiper, G.G.J.M.; Carlsson, B.; Grandien, K.; Enmark, E.; Häggblad, J.; Nilsson, S.; Gustafsson, J-A. *Endocrinology* **1997**, 138, 863-870.

† The RBA and K_i values tabulated here are for the ERα as defined in the reference. "The RBA of each competitor is calculated as the ratio of concentrations of E2 and competitor required to reduce the specified radioligand binding by 50% (= ratio of IC_{50} values). The RBA value of E2 was arbitrarily set at 100."

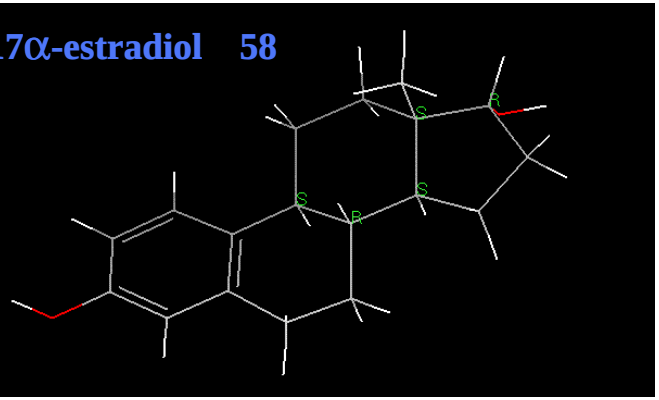
‡ The authors referenced above used the Cheng-Prusoff formula to calculate K_i values from their IC_{50} results.

Some of the Ligands Studied

Diethylstilbestrol 468

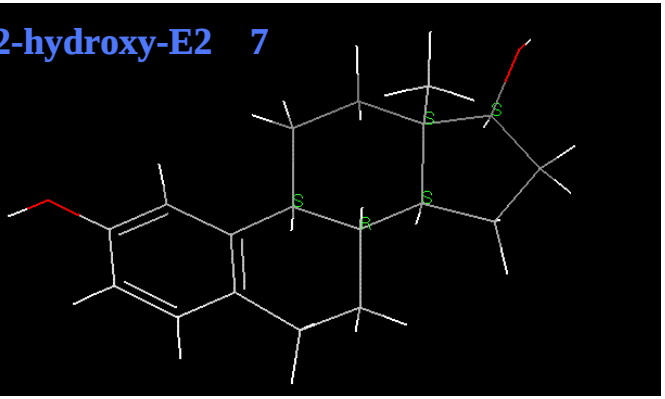


17α-estradiol 58

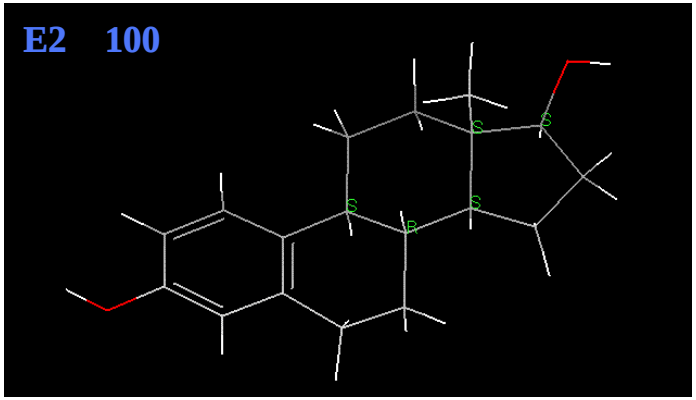


4-hydroxy-tamoxifen 178

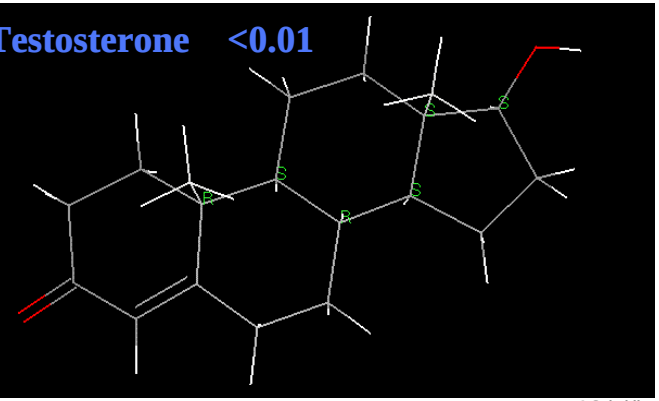
2-hydroxy-E2 7



E2 100

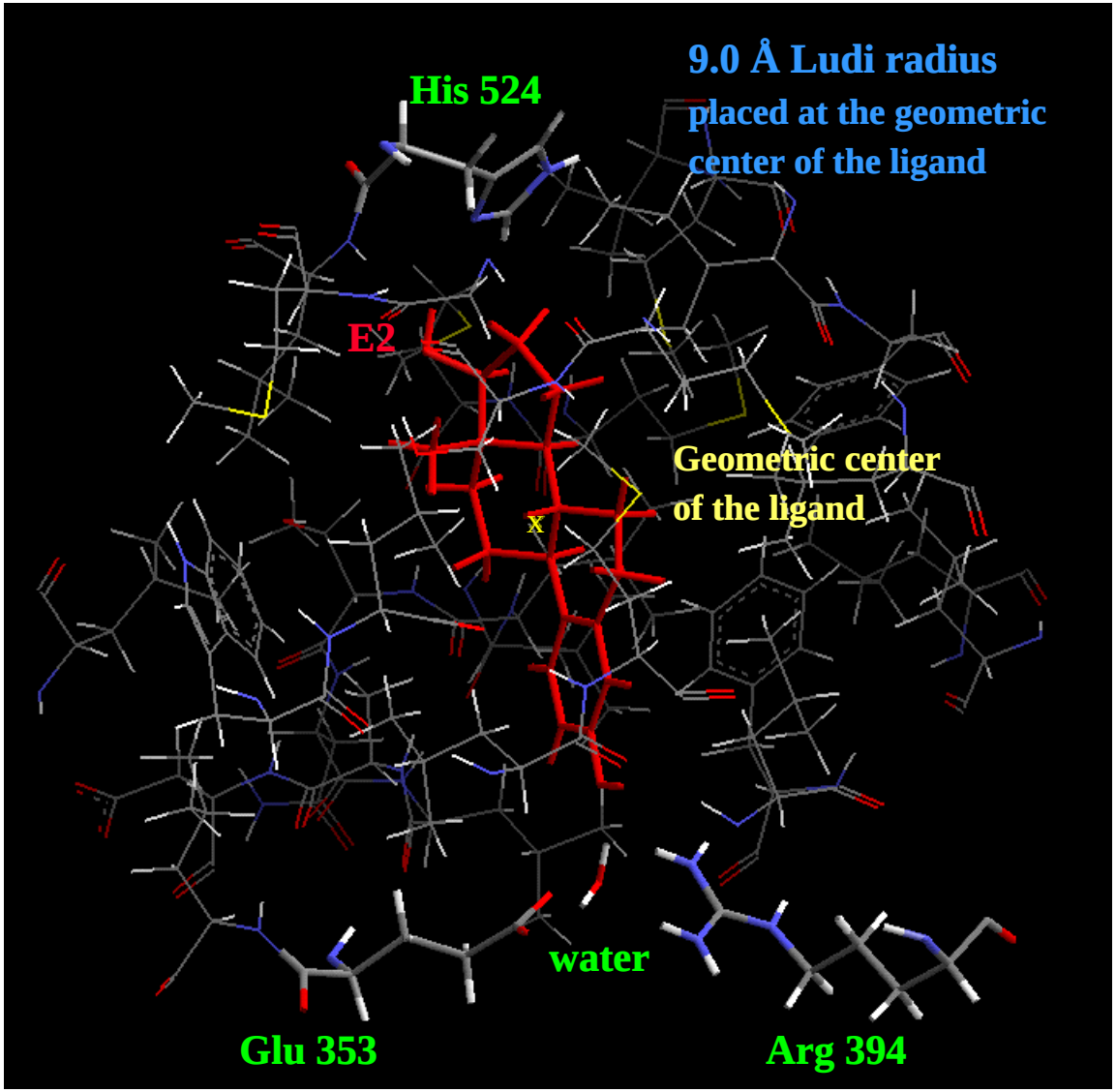


Testosterone <0.01

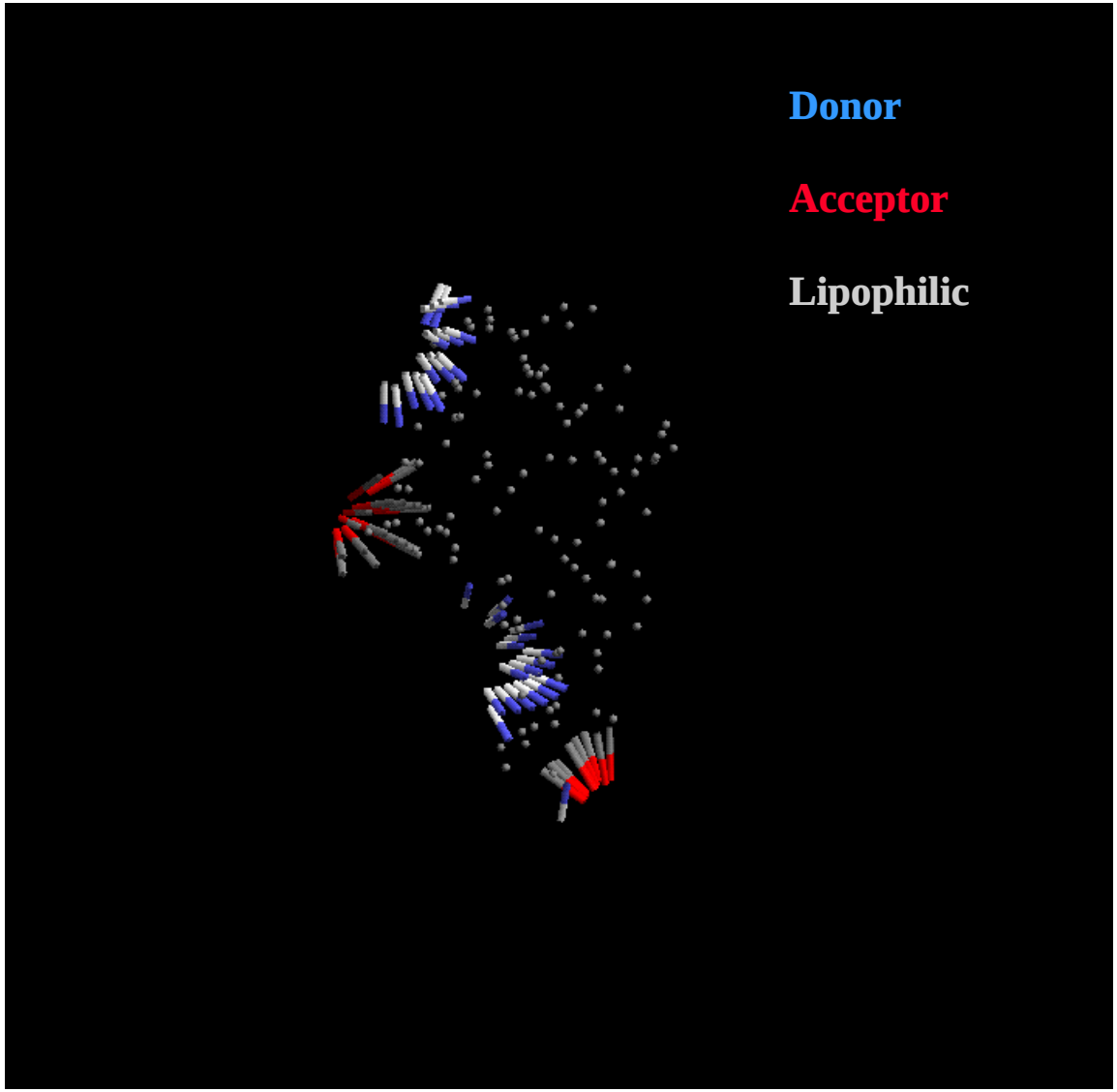


A Subsidiary of Pharmacopeia Inc.

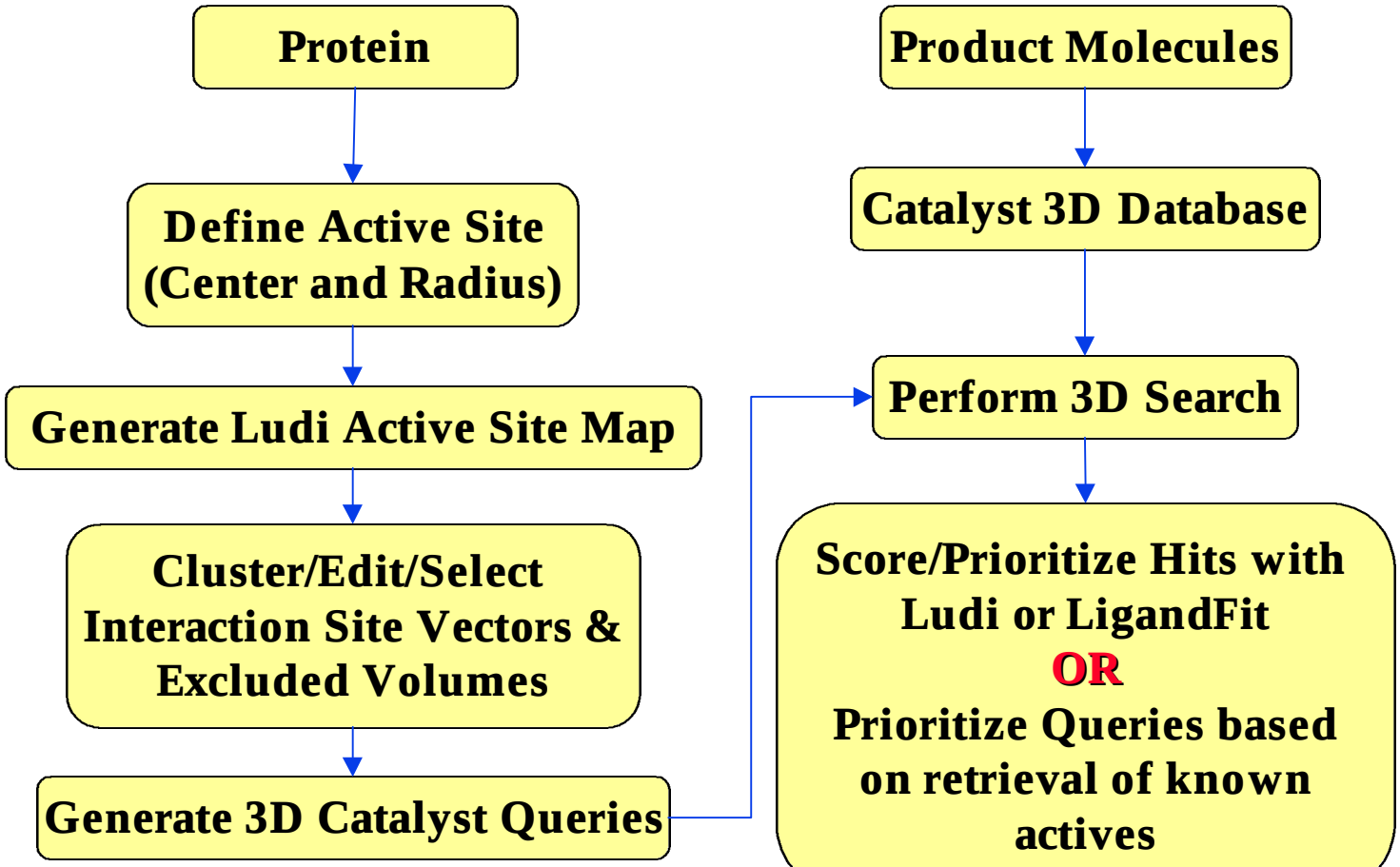
Model for the Binding Site



Interaction Model for the Binding Site

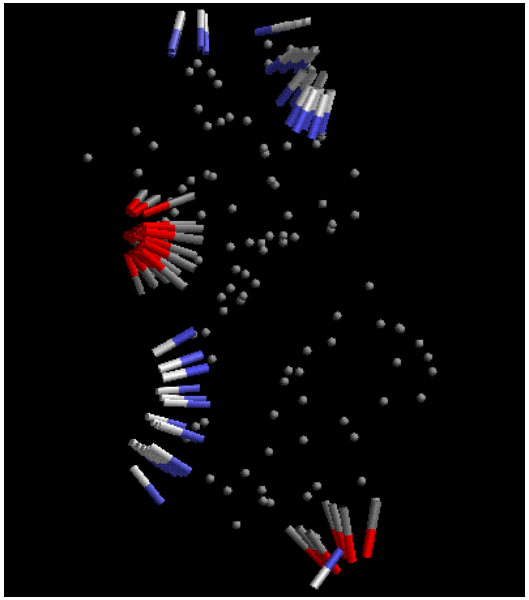


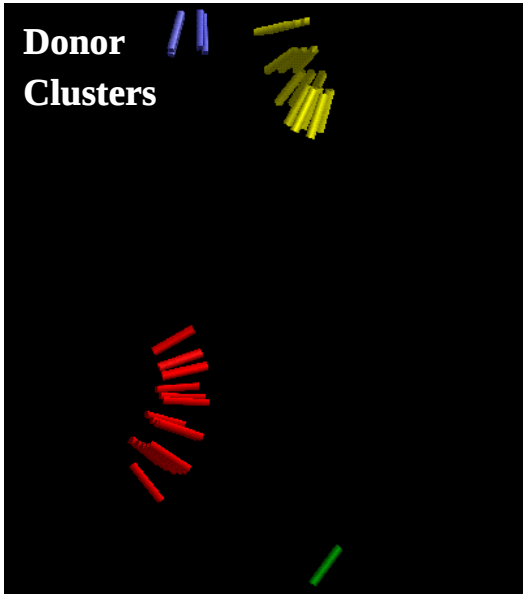
Cerius^2 Structure Based-Focusing

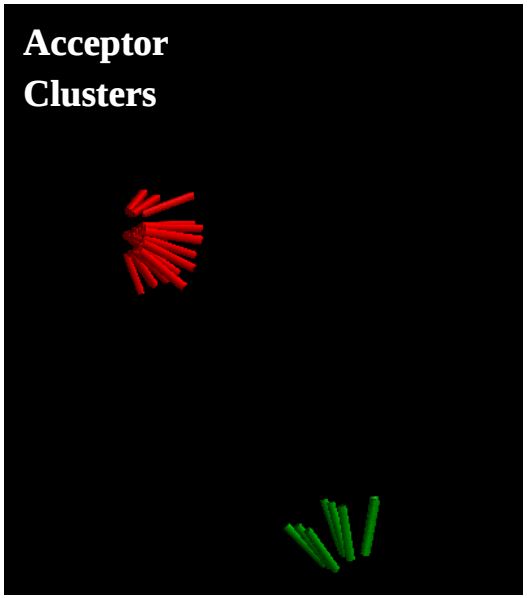


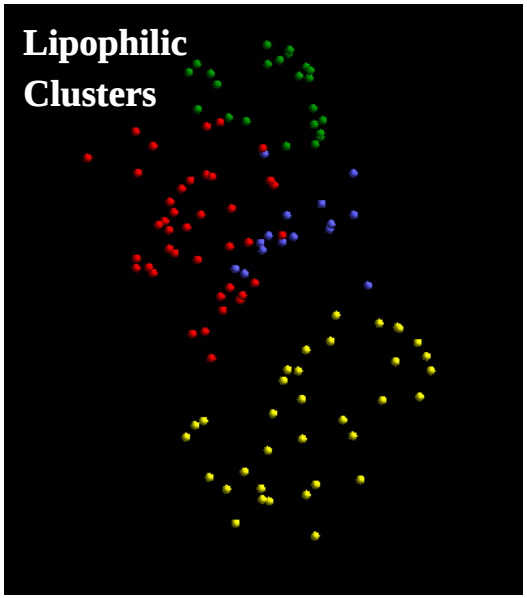
A Subsidiary of Pharmacopeia Inc.

Clustering of the Interaction Model









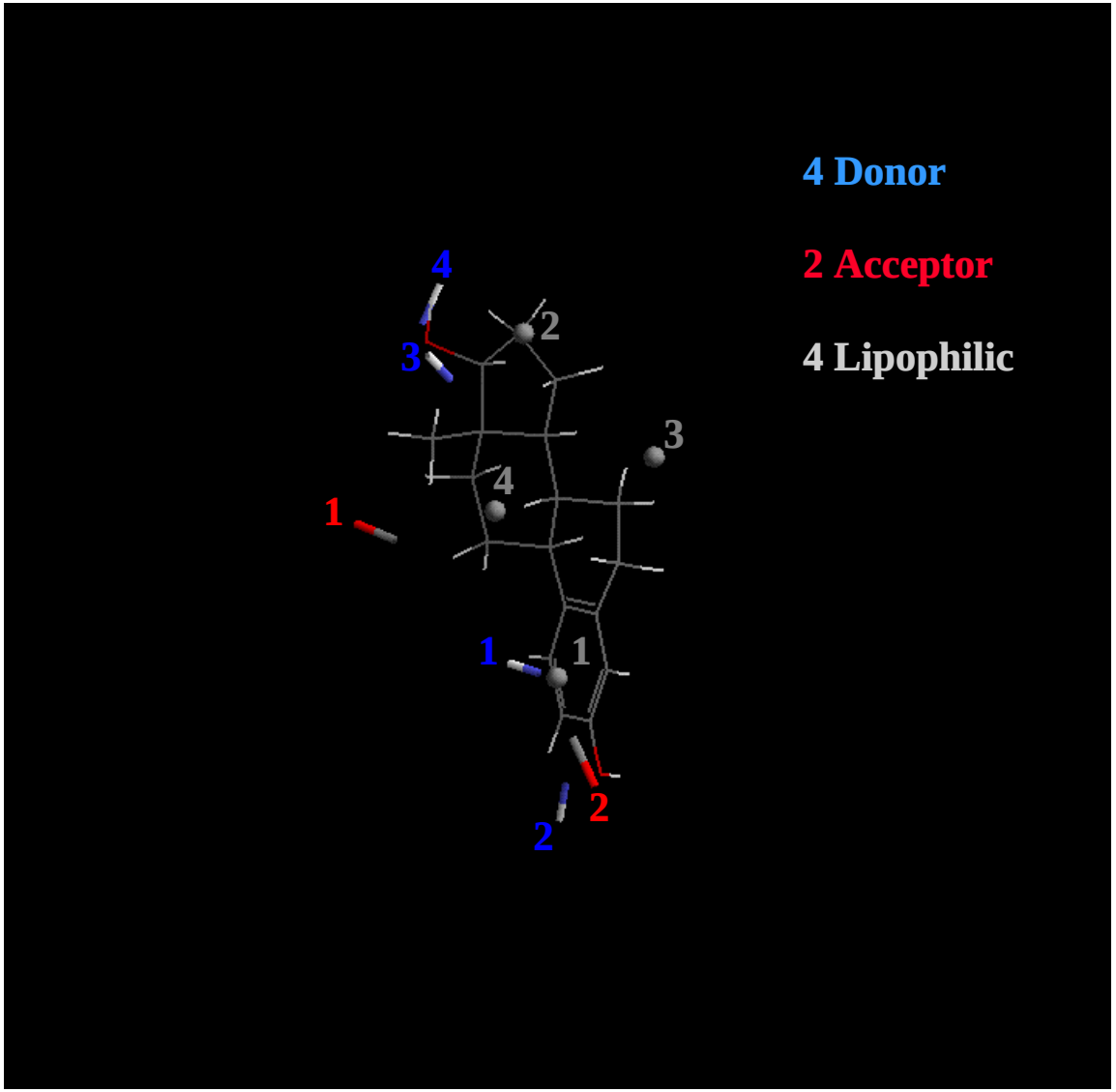
Model orientation is slightly rotated compared to the previous slide

Features were clustered using the Complete Linkage Method and specifying for the number of clusters to be formed as

4 Donor
2 Acceptor
4 Lipophilic

Some editing of the clusters was done by hand

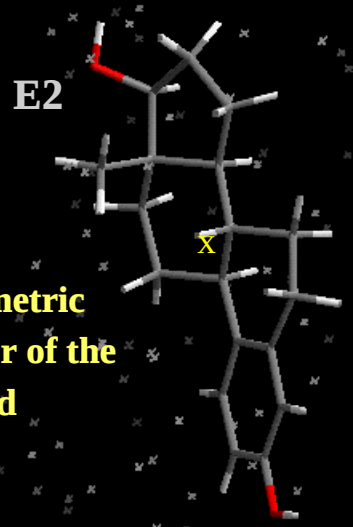
Cluster Centers of Interaction Model Overlaid on the E2 Ligand



A Subsidiary of Pharmacopeia Inc.

Excluded Regions Occupied by the Receptor

Find Closest Atoms option for heavy atoms within 11.0 Å of the geometric center of the ligand (228 pts)



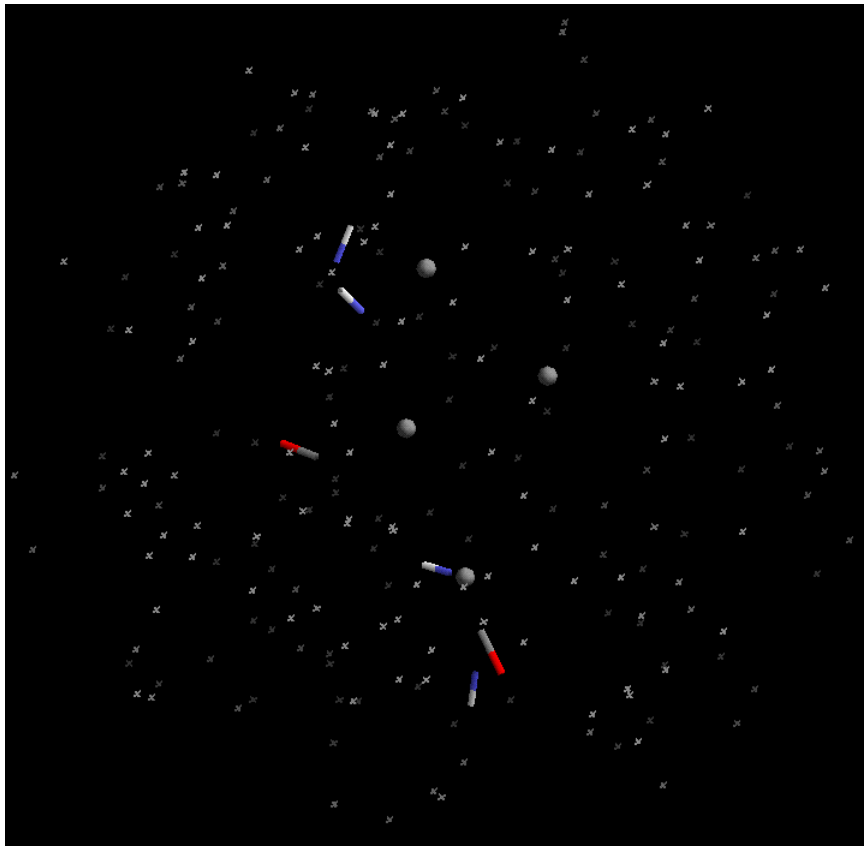
Geometric center of the ligand

Radii for the excluded volumes ranging from 1.0 to 2.0 Å were tried with and without hydrogen atoms

For this study the best results were obtained using a 1.3 Å radius for the excluded volumes of the heavy atoms (no hydrogens)

Generation of the Queries

Queries were generated using the excluded regions and combinations of the available features. The only restriction used in generating these queries was a minimum distance limit of 1.5 Å between features.



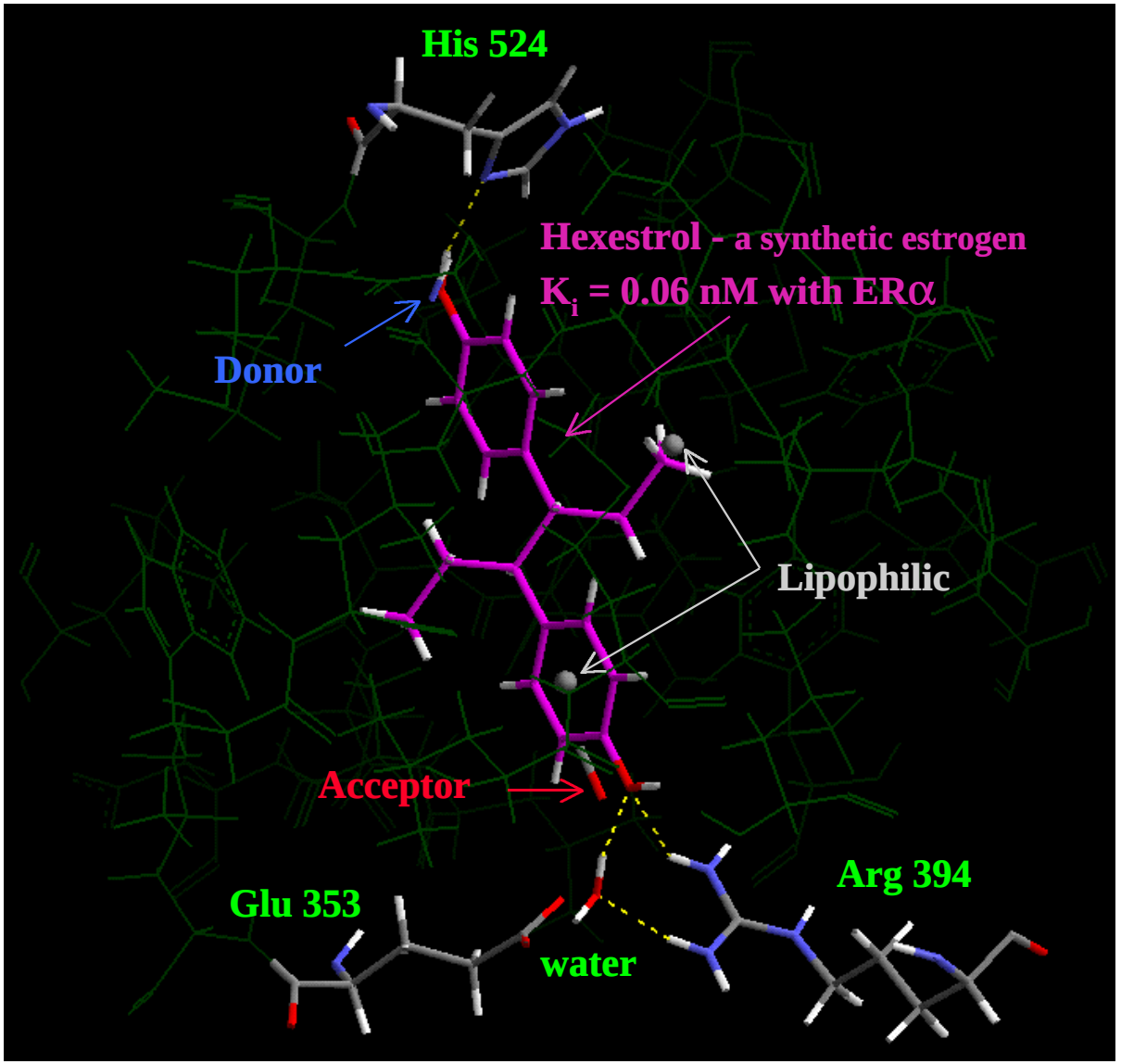
# of Features in each Query	# of Queries Generated	% of Queries Producing Hits
3 Feature	112	57
4 Feature	182	33
5 Feature	196	13
6 Feature	140	3

$$\text{Maximum No. of Queries} = \frac{N!}{n!\,(N-n)!}$$

N = Number of Features Available

n = Number of Features in a Query

Docking to a 4 Feature Query of the Estrogen Receptor



Tabulation of the Search Results

Hit file handle corresponds to query file (*.chm)

A value of "1" indicates hits for 1 or more conformations of the ligand

Ligands with Decreasing RBA →

Hits File MDL SD Format	Donor	Acceptor	Features Comprising the Query	Lipophilic	Diethylstilbestrol	Hexestrol	Dienestrol	4-hydroxy-tamoxifen	E2	Coumestrol	ICI164384	Estrone	17α-estradiol	Nafoxidene	Moxestrol	Clomifene	β-zearalanol	Estriol	4-hydroxy-E2	2-hydroxy-E2	Tamoxifen	5-androstanediol	Genistein	3β-androstanediol	Norethynodrel	4-androstenediol	3α-androstanediol	Norethindrone	BisphenolA	5α-dihydrotestosterone	Dehydroepiandosterone	Methoxyclor	Testosterone	Progesterone	Represented
feat3_5.sd	1,2			2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
feat3_7.sd	1,2			4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
feat3_11.sd	1,3			1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
feat3_12.sd	1,3			2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
feat3_13.sd	1,3			3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
feat3_17.sd	1,4			1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
feat3_23.sd	1	1		2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
feat3_24.sd	1	1		3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
feat3_29.sd	1	2		4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
feat3_30.sd	1			1,2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2
.																																			
.																																			
.																																			
feat3_111.sd				1,3,4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2
feat3_112.sd				2,3,4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
3 Feature Totals					5	20	5	0	13	4	0	2	6	0	1	0	0	12	13	8	0	11	10	13	0	11	6	2	27	2	2	0	1	0	

Indicates the number of queries which produced hits for a particular ligand

Indicates the number of ligands which had hits for a particular query

Steroids, synthetic estrogens, phytoestrogens, industrial chemicals, antiestrogens

Examples of Hits Obtained for Individual Queries

Ligands with Decreasing RBA →

Hits File MDL SD Format	Donor	Features Comprising the Query: Acceptor	Lipophilic	Diethylstilbestrol	Hexestrol	Dienestrol	4-hydroxy-tamoxifen	E2	Coumestrol	ICI164384	Estrone	17α-estradiol	Nafoxidene	Moxestrol	Clomifene	β-zearalanol	Estriol	4-hydroxy-E2	2-hydroxy-E2	Tamoxifen	5-androstanediol	Genistein	3β-androstanediol	Norethynodrel	4-androstenediol	3α-androstanediol	Norethindrone	BisphenolA	5α-dihydrotestosterone	Dehydroepiandosterone	Methoxychlor	Testosterone	Progesterone	Represented
feat3_55.sd	2		2,4	1	1	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	6
feat3_85.sd	4	2	3	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
feat3_88.sd	4		1,3	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	3
feat3_89.sd	4		1,4	0	1	0	0	1	0	0	0	1	0	0	0	0	1	0	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	8
3 Feature Totals				5	20	5	0	13	4	0	2	6	0	1	0	0	12	13	8	0	11	10	13	0	11	6	2	27	2	2	0	1	0	
feat4_112.sd	2		2,3,4	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
feat4_159.sd	4	2	1,3	1	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
feat4_166.sd	4		1,3,4	1	1	0	0	1	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
feat4_181.sd		2	2,3,4	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
4 Feature Totals				5	21	4	0	15	1	0	2	7	0	0	0	0	11	11	7	0	6	3	8	1	7	6	0	18	0	2	0	0	0	
feat5_100.sd	1	1	2,3,4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
feat5_134.sd	2,4		1,3,4	1	1	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	7
feat5_188.sd	4	2	1,3,4	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
feat5_195.sd		1	1,2,3,4	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
5 Feature Totals				1	11	3	0	6	0	0	1	2	0	0	0	0	3	3	0	0	1	0	2	0	2	2	0	5	0	0	0	0	0	
feat6_90.sd	1	1	1,2,3,4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
feat6_106.sd	2,3		1,2,3,4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
feat6_132.sd	3	1	1,2,3,4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
feat6_138.sd	4	1	1,2,3,4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
6 Feature Totals				0	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Steroids, synthetic estrogens, phytoestrogens, industrial chemicals, antiestrogens

How to Analyze Hit Lists?

Different metrics can be used to evaluate the quality of a hit list

- Percent yield *(%Y)*: the percentage of the known actives in the hit list

$$\%Y = \frac{H_a}{H_t} \times 100$$

- Percent of actives *(%A)*: the percentage of known active compounds retrieved from the database

$$\%A = \frac{H_a}{A} \times 100$$

$\boldsymbol{H_a}$ is the number of active ligands which were hit

$\boldsymbol{H_t}$ is the total number of hits obtained

$\boldsymbol{A}$ is the number of active ligands in the database

How to Analyze Hit Lists?

Different metrics can be used to evaluate the quality of a hit list

- Enrichment (*E*): indicates how many time more richer the hit list is than the original database with respect to the yield of actives

$$E = \left(\frac{H_a / H_t}{A / D} \right) = \frac{H_a \times D}{H_t \times A}$$

H_a is the number of active ligands which were hit

H_t is the total number of hits obtained

A is the number of active ligands in the database

D is the number of compounds in the database

How to Analyze Hit Lists?

Different metrics can be used to evaluate the quality of a hit list

- Goodness of hit list (*GH*):[1] a weighted linear combination of *%Y* and *%A*, with correction to address database and hit list size differences

$$GH = \left(\frac{H_a (3A + H_t)}{4 H_t A} \right) \times \left(1 - \frac{H_t - H_a}{D - A} \right)$$

H_a is the number of active ligands which were hit

H_t is the total number of hits obtained

A is the number of active ligands in the database

D is the number of compounds in the database

[1] **Güner, O. F. and Henry, D. R. in "Pharmacophore Perception, Development, and Use in Drug Design," 1999, in press.**

Hit Totals for the Queries

(antiestrogens not included)

	More Active							Less Active																
Queries	Diethylstilbestrol	Hexestrol	Dienestrol	E2	Coumestrol	Estrone	17α-estradiol	Moxestrol	β-zearalanol	Estriol	4-hydroxy-E2	2-hydroxy-E2	5-androstanediol	Genistein	3β-androstanediol	Norethynodrel	4-androstenediol	3α-androstanediol	Norethindrone	BisphenolA	5α-dihydrotestosterone	Dehydroepiandosterone	Methoxychlor	Testosterone	Progesterone
						60	58	43	16	14	13	7			3	0.7									
3 Feature	5	20	5	13	4	2	6	1	0	12	13	8	11	10	13	0	11	6	2	27	2	2	0	1	0
4 Feature	5	21	4	15	1	2	7	0	0	11	11	7	6	3	8	1	7	6	0	18	0	2	0	0	0
5 Feature	1	11	3	6	0	1	2	0	0	3	3	0	1	0	2	0	2	2	0	5	0	0	0	0	0
6 Feature	0	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Queries	Percentage of Actives Hit	Percentage of Less Actives Hit	E	GH
3 Feature	100	78	1.2	0.11
4 Feature	100	61	1.4	0.21
5 Feature	86	39	1.6	0.34
6 Feature	29	0	3.6	0.82

Hit Totals for the Queries

(antiestrogens not included)

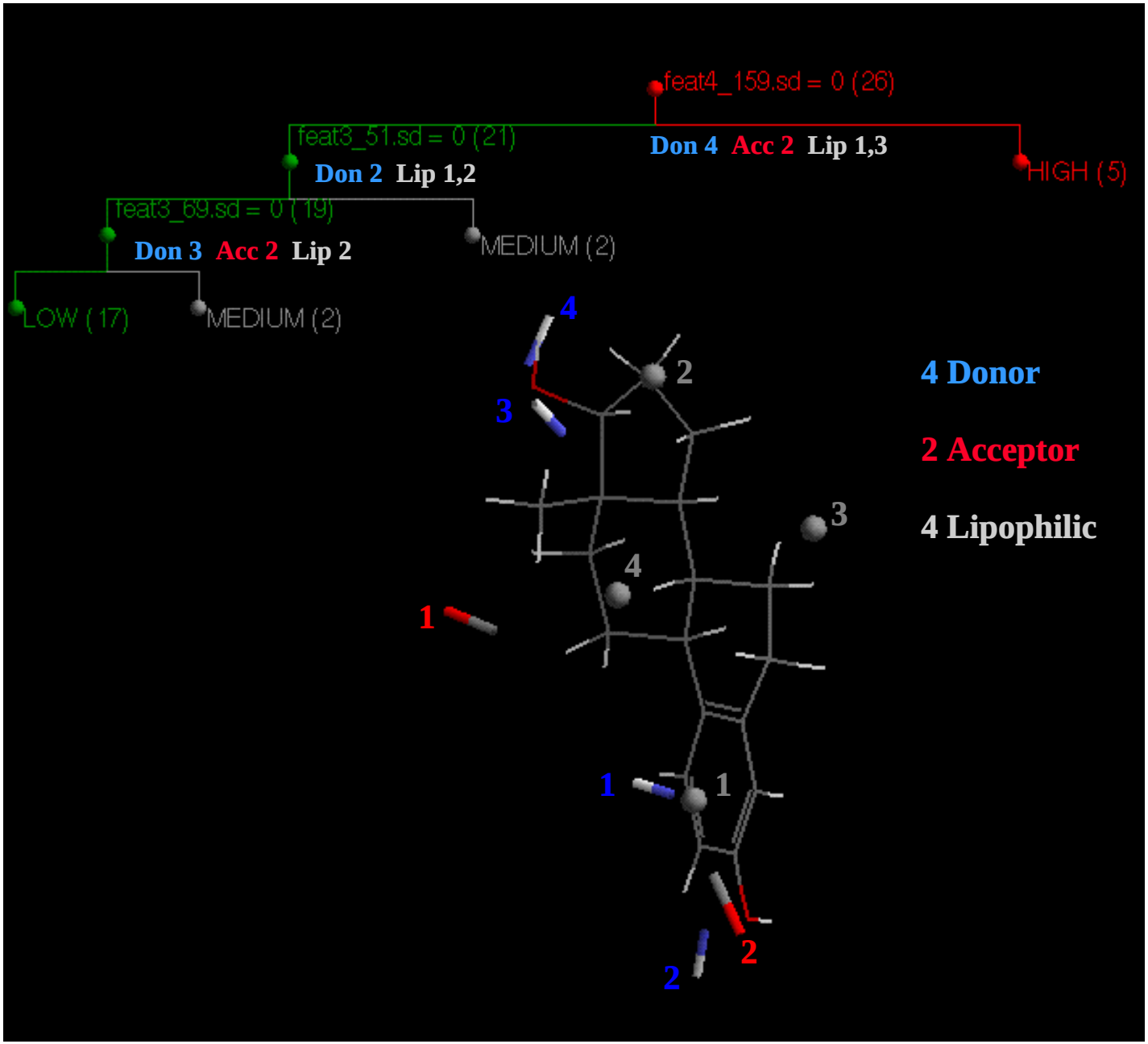
	More Active															Less Active									
Queries	Diethylstilbestrol	Hexestrol	Dienestrol	E2	Coumestrol	Estrone	17α-estradiol	Moxestrol	β-zearalanol	Estriol	4-hydroxy-E2	2-hydroxy-E2	5-androstanediol	Genistein	3β-androstanediol	Norethynodrel	4-androstenediol	3α-androstanediol	Norethindrone	BisphenolA	5α-dihydrotestosterc	Dehydroepiandoster	Methoxychlor	Testosterone	Progesterone
					60	58	43	16	14	13	7	3	0.7												
3 Feature	5	20	5	13	4	2	6	1	0	12	13	8	11	10	13	0	11	6	2	27	2	2	0	1	0
4 Feature	5	21	4	15	1	2	7	0	0	11	11	7	6	3	8	1	7	6	0	18	0	2	0	0	0
5 Feature	1	11	3	6	0	1	2	0	0	3	3	0	1	0	2	0	2	2	0	5	0	0	0	0	0
6 Feature	0	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Queries	Percentage of Actives Hit	Percentage of Less Actives Hit	E	GH
3 Feature	93	70	1.1	0.22
4 Feature	87	50	1.2	0.38
5 Feature	67	30	1.3	0.52
6 Feature	13	0	1.7	0.78

Conclusions

This is a very challenging system for this type of study. Despite the difficulties with this system, SBF

- Provided distinction between the agonist and antagonist ligands examined.
- Generated queries which produced hits for specific ligands. Queries producing the desired hits could be used to search large databases.
- Provide information which could be used to further refine the queries.
- Generated queries which as a whole showed enrichment of the more active ligands examined.



Analysis of Hits from Queries using Recursive Partitioning

If (Don 4 & Acc 2 & Lip 1, 3) then
HIGHLY ACTIVE

else if (Don 2 & Lip 1,2) then
MEDIUM ACTIVE

else if (Don 3 & Acc 2 & Lip 2) then
MEDIUM ACTIVE
else
WEAK ACTIVE
end if

Thanks to...

Paul Kirchhoff
Jeff Jiang
Rob Brown
Marvin Waldman
Scott Kahn