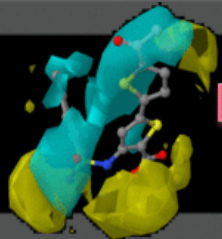


Ligand-Based Drug Design



SAPIENZA
UNIVERSITÀ DI ROMA



	Coefficient	Symbol	Equation	Limits
1	Squared Correlation Coefficient	R^2 or r^2	$r^2 = 1 - \frac{\sum_{i=1}^N (Y_{\text{exp},i} - Y_{\text{calc},i})^2}{\sum_{i=1}^N (Y_{\text{exp},i} - \bar{Y})^2}$	$0 \leq r^2 \leq 1$
2	Cross-Validated R^2	Q^2 or q^2		$q^2 = 1 - \frac{\sum_{i=1}^N (Y_{\text{exp},i} - Y_{\text{pred},i})^2}{\sum_{i=1}^N (Y_{\text{exp},i} - \bar{Y})^2}$
3	Y-scrambling	r^2_{ys} and q^2_{ys}		$r^2_{\text{ys}} \leq r^2$ $q^2_{\text{ys}} \leq q^2$
4	Prediction Error	<i>SDEP</i>	$SDEP = \sqrt{\frac{\sum_{i=1}^N (Y_{\text{exp},i} - Y_{\text{pred},i})^2}{N}}$	As low as possible

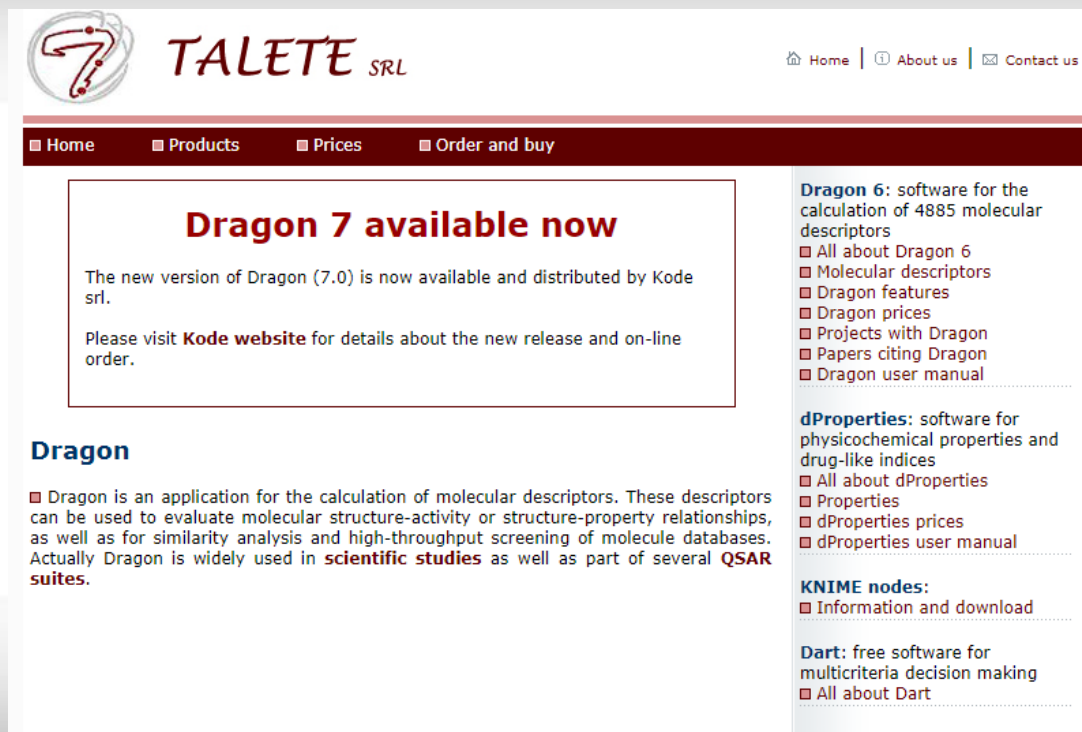


Descriptors for QSAR

Molecular descriptors are numerical values that characterize properties of molecules

The descriptors fall into Four classes

- Topological
- Geometrical
- Electronic
- Hybrid or 3-D Descriptors
- Fingerprints



TALETE SRL

Home | About us | Contact us

Home | Products | Prices | Order and buy

Dragon 7 available now

The new version of Dragon (7.0) is now available and distributed by Kode srl.

Please visit **Kode website** for details about the new release and on-line order.

Dragon

Dragon is an application for the calculation of molecular descriptors. These descriptors can be used to evaluate molecular structure-activity or structure-property relationships, as well as for similarity analysis and high-throughput screening of molecule databases. Actually Dragon is widely used in **scientific studies** as well as part of several **QSAR suites**.

Dragon 6:

software for the calculation of 4885 molecular descriptors

- All about Dragon 6
- Molecular descriptors
- Dragon features
- Dragon prices
- Projects with Dragon
- Papers citing Dragon
- Dragon user manual

dProperties:

software for physicochemical properties and drug-like indices

- All about dProperties
- Properties
- dProperties prices
- dProperties user manual

KNIME nodes:

- Information and download

Dart:

free software for multicriteria decision making

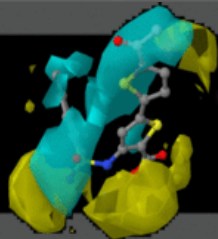
- All about Dart



Descriptors for QSAR

Dragon 7.0 calculates **5,270** molecular descriptors, organized in different **logical blocks** as in the previous versions. Blocks are further divided into sub-blocks to make management, selection, and analysis of descriptors easier. Following, the summary of molecular descriptors blocks calculated by Dragon 7.0 is reported. The [complete list](#) of all 5,270 calculated descriptor is also available.

Block no.	Block name	# Descriptors	Block no.	Block name	# Descriptors
1	Constitutional	47	16	RDF descriptors	210
2	Ring descriptors	32	17	3D-MoRSE descriptors	224
3	Topological indices	75	18	WHIM descriptors	114
4	Walk and path counts	46	19	GETAWAY descriptors	273
5	Connectivity indices	37	20	Randic molecular profiles	41
6	Information indices	50	21	Functional groups count	154
7	2D matrix-based descriptors	607	22	Atom-centred fragments	115
8	2D autocorrelations	213	23	Atom-type E-state indices	172
9	Burden eigenvalues	96	24	CATS 2D	150
10	P-VSA-like descriptors	55	25	2D Atom Pairs	1596
11	ETA indices	23	26	3D Atom Pairs	36
12	Edge adjacency indices	324	27	Charge descriptors	15
13	Geometrical descriptors	38	28	Molecular properties	20
14	3D matrix-based descriptors	99	29	Drug-like indices	28
15	3D autocorrelations	80	30	CATS 3D	300



1980: The Goodford's GRID (Toward 3-D QSAR)

J. Med. Chem. 1985, 28, 849–857

849

Articles

A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules

P. J. Goodford

*The Laboratory of Molecular Biophysics, The Rex Richards Building, University of Oxford, Oxford OX1 3QU, England.
Received August 3, 1984*

The interaction of a probe group with a protein of known structure is computed at sample positions throughout and around the macromolecule, giving an array of energy values. The probes include water, the methyl group, amine nitrogen, carboxy oxygen, and hydroxyl. Contour surfaces at appropriate energy levels are calculated for each probe and displayed by computer graphics together with the protein structure. Contours at negative energy levels delineate regions of attraction between probe and protein and are found at known ligand binding clefts in particular. The contours also enable other regions of attraction to be identified and facilitate the interpretation of protein–ligand energetics. They may, therefore, be of value for drug design.

Description of Molecules with Molecular Interaction Fields (MIF)



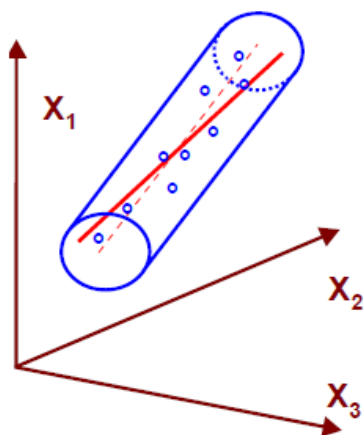
PCA

<http://setosa.io/ev/principal-component-analysis/>



197x-198x: The Wold's PLS (Toward 3-D QSAR)

PLS (Partial Least Squares) Analysis



From $u = kt$

($k = \text{constants } k_1, k_2 \dots k_j$;
 $j = \text{number of PLS vektors}$)

follows:

$$BA_i = a_1 S_{i1} + a_2 S_{i2} + a_3 S_{i3} + \dots + a_m S_{im} + b_1 E_{i1} + b_2 E_{i2} + b_3 E_{i3} + \dots + b_m E_{im}$$

Y vector

X matrix

BA_1	S_{11}	S_{12}	S_{13}	S_{14}	\dots	S_{1m}	E_{11}	E_{12}	\dots	E_{1m}
BA_2	S_{21}	S_{22}	S_{23}	S_{24}	\dots	S_{2m}	E_{21}	E_{22}	\dots	E_{2m}
BA_3	S_{31}	S_{32}	S_{33}	S_{34}	\dots	S_{3m}	E_{31}	E_{32}	\dots	E_{3m}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
BA_n	S_{n1}	S_{n2}	S_{n3}	S_{n4}	\dots	S_{nm}	E_{n1}	E_{n2}	\dots	E_{nm}

PLS analysis

SAMPLS analysis

u_{11}
u_{12}
u_{13}
\vdots
u_{1n}

t_{11}	t_{21}
t_{12}	t_{22}
t_{13}	t_{23}
\vdots	\vdots
t_{1n}	t_{2n}

c_{11}	c_{12}	c_{13}	\dots	c_{1n}
c_{21}	c_{22}	c_{23}	\dots	c_{2n}
c_{31}	c_{32}	c_{33}	\dots	c_{3n}
\vdots	\vdots	\vdots	\vdots	\vdots
c_{n1}	c_{n2}	c_{n3}	\dots	c_{nn}

PLS vectors
(latent variables)

covariance matrix

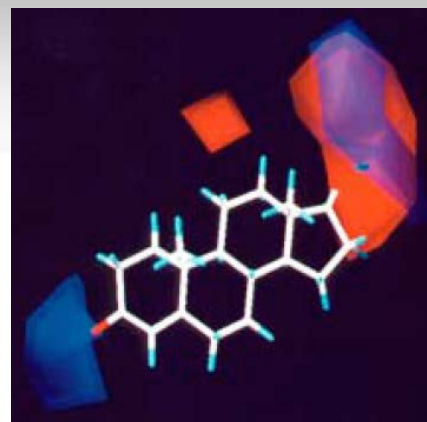
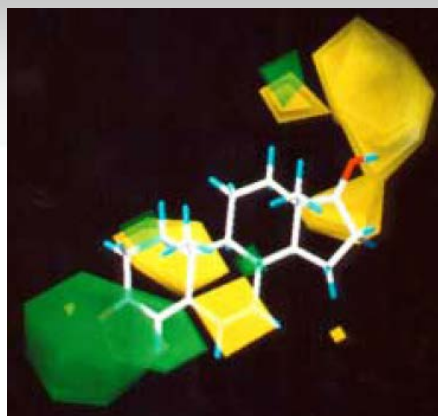
Reduction of Dimensionality into Few New Highly Informative Entities

----- Principal Components -----



1988: The First 3-D QSAR

PLS + MIF → CoMFA!



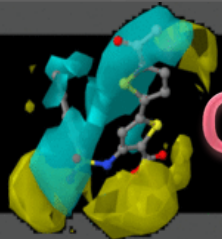
J. Am. Chem. Soc. **1988**, *110*, 5959–5967

5959

Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins

Richard D. Cramer, III,* David E. Patterson, and Jeffrey D. Bunce

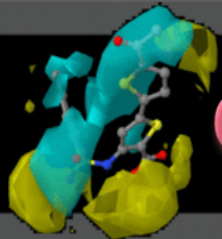
*Contribution from Tripos Associates, 1699 South Hanley Road,
St. Louis, Missouri 63144. Received January 5, 1988*



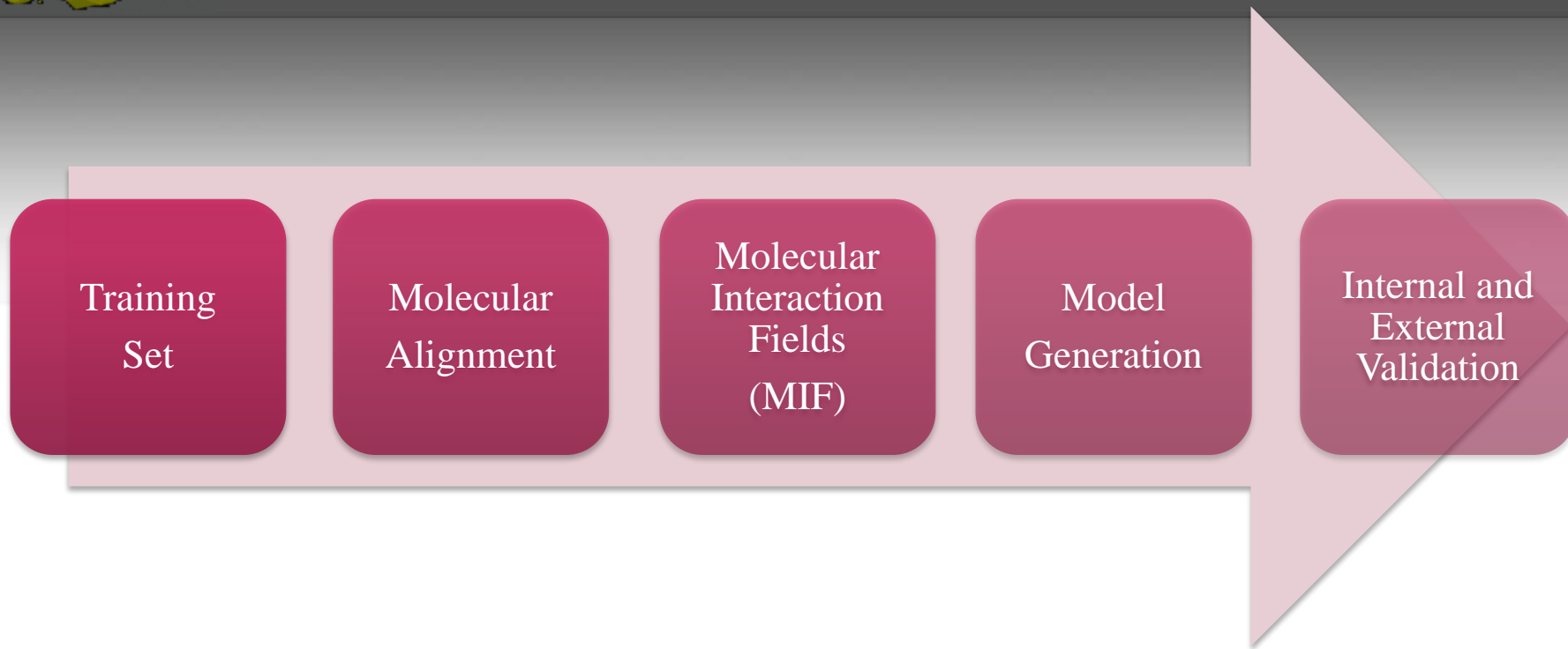
- Physical properties are measured for the molecule as a whole
- Properties are calculated using computer software
- No experimental constants or measurements are involved
- Properties are known as 'Fields'
- Steric field - defines the size and shape of the molecule
- Electrostatic field - defines electron rich/poor regions of molecule
- Hydrophobic properties are relatively unimportant

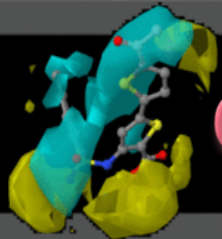
Advantages over classical QSAR

- No reliance on experimental values (*i.e.* logP)
- Can be applied to molecules with unusual substituents
- Not restricted to molecules of the same structural class
- Improved predictive capability



CoMFA/3-D QSAR Procedure





CoMFA/3-D QSAR Procedure

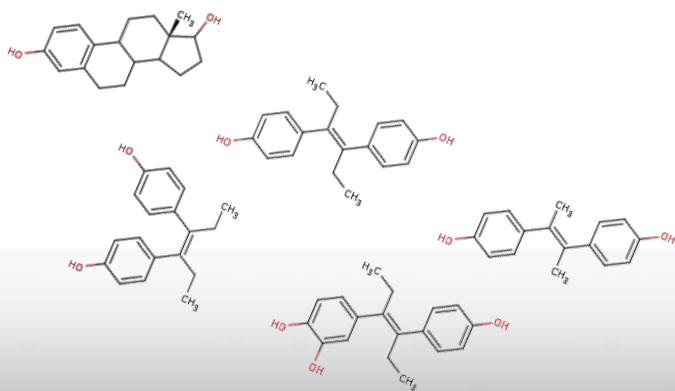
**Training
Set**

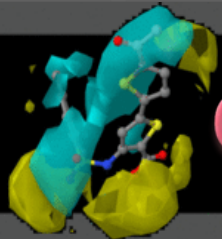
Molecular
ligment

Molecular
Interaction
Fields
(MIF)

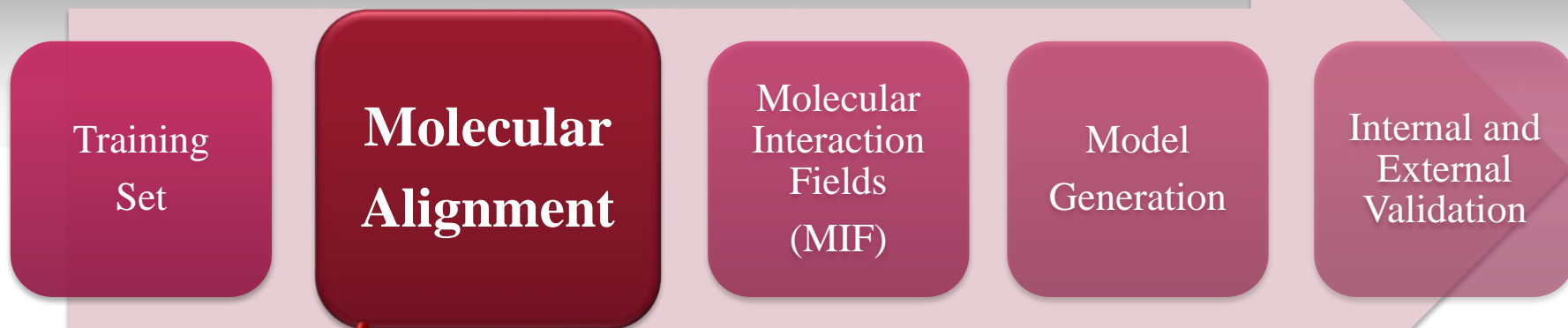
Model
Generation

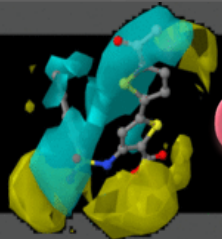
Internal and
External
Validation



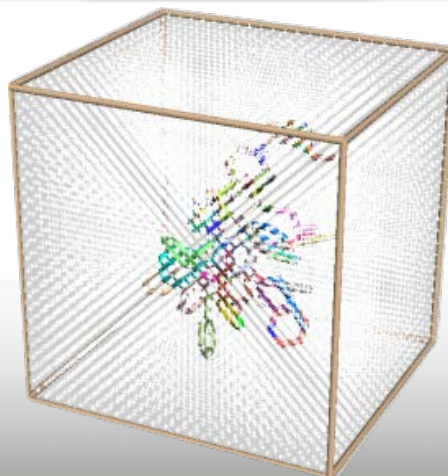
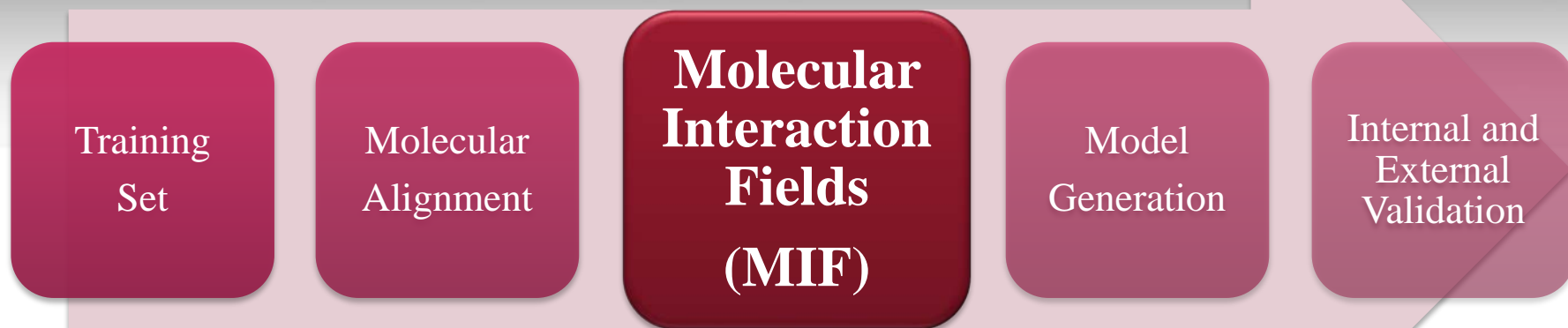


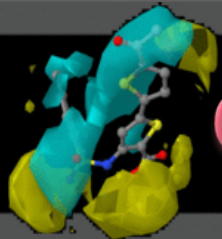
CoMFA/3-D QSAR Procedure



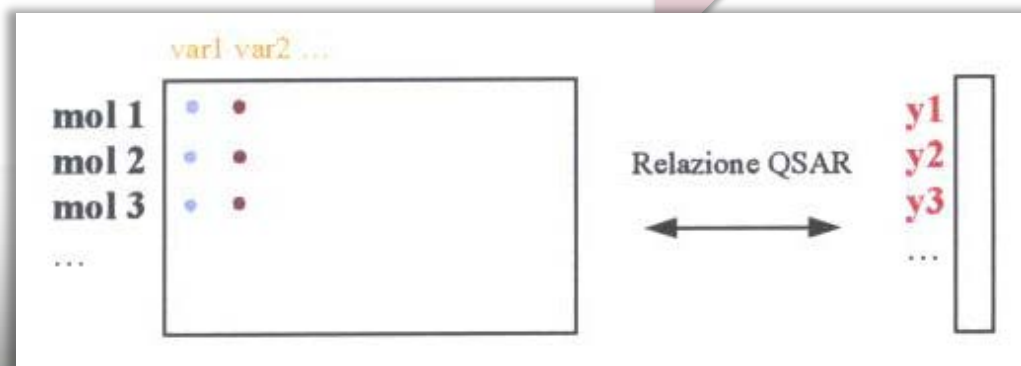
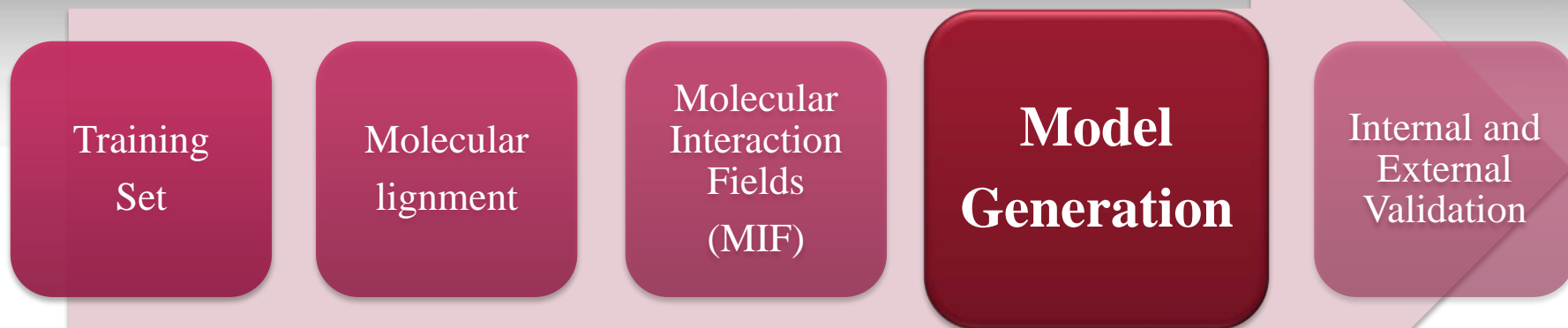


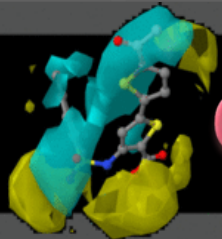
CoMFA/3-D QSAR Procedure



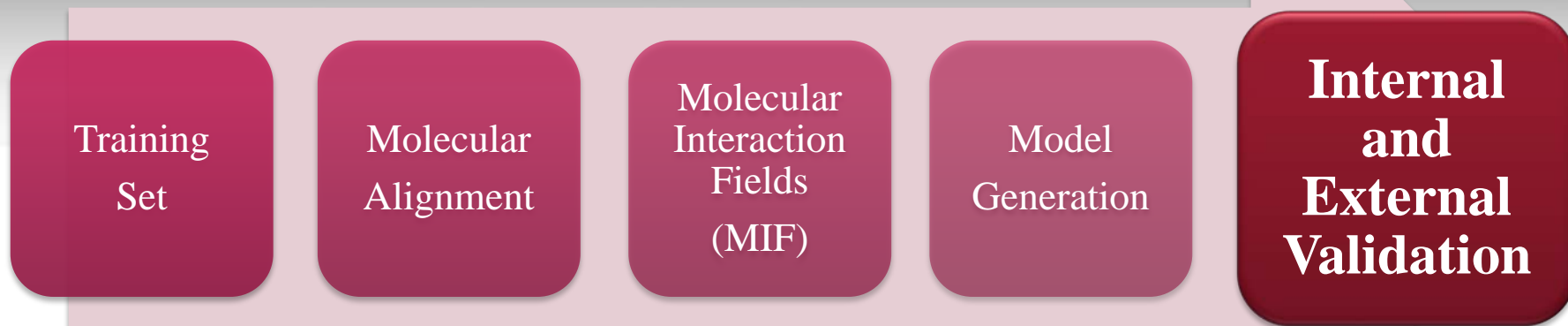


CoMFA/3-D QSAR Procedure

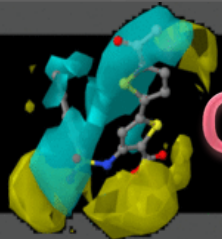




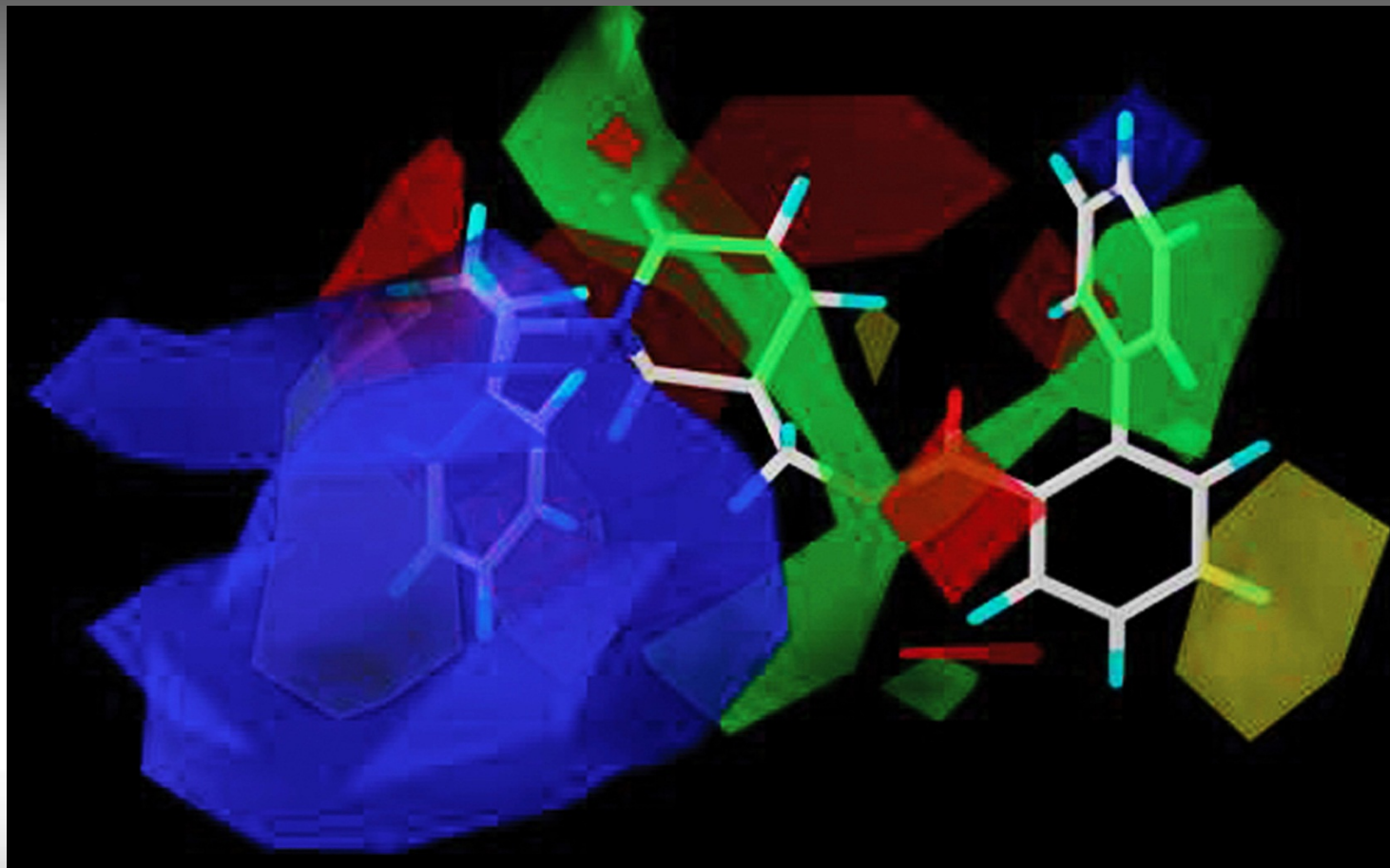
CoMFA/3-D QSAR Procedure



$$SDEP = \sqrt{\sum (y_{\text{exp}} - y_{\text{pred}})^2 / n - 1}$$
$$q^2 = 1 - \sum (y_{\text{exp}} - y_{\text{pred}})^2 / \sum (y_{\text{exp}} - y)^2$$



CoMFA/3-D QSAR Procedure





CoMFA Limits

CRUCIAL POINTS!

- **Conformation of the training set molecules**
- **Superimposition of the training set molecules (molecular alignment rules)**

Original CoMFA Tricks:

- **Very rigid molecules: steroid scaffold!!!!**
- **Directed superimposed atom by atom!!!!**



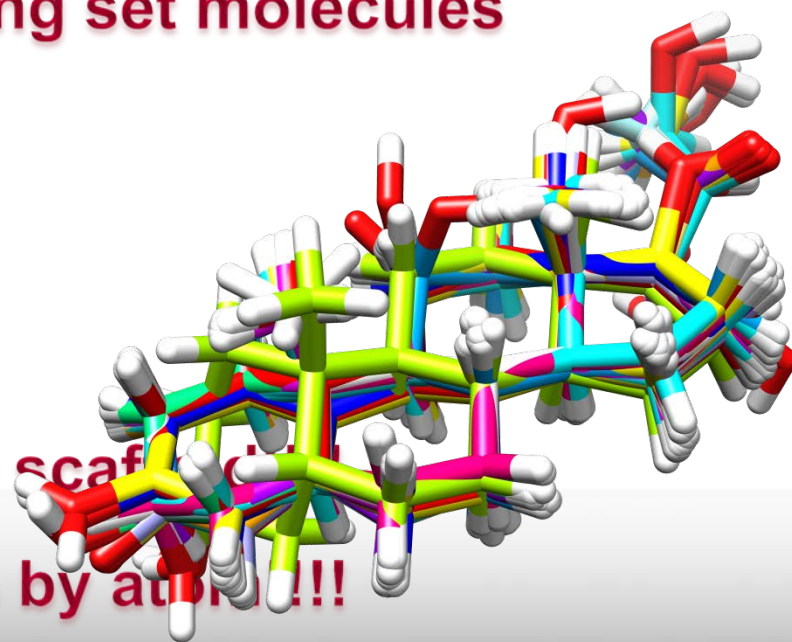
CoMFA Limits

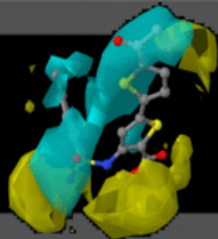
CRUCIAL POINTS!

- Conformation of the training set molecules
- Superimposition of the training set molecules (molecular alignment rules)

Original CoMFA Tricks:

- Very rigid molecules: steroid scaffolds
- Directed superimposed atom by atom!!!





1991: The WO CoMFA Patent and 3-D QSAR Development

The Lattice Model: A General Paradigm for Shape-Related Structure/Activity Correlation

Cramer, R.D., and Milne, M., *Abstracts ACS Meeting*, Honolulu, 1979, COMP 44.

SIAM J. Sci. and Stat. Comput. / Volume 5 / Issue 3

The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses

S. Wold, A. Ruhe, H. Wold, and W. J. Dunn, III

SIAM J. Sci. and Stat. Comput. Volume 5, Issue 3, pp. 735-743

J. Am. Chem. Soc. 1988, 110, 5959-5967

5959

Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins

Richard D. Cramer, III,* David E. Patterson, and Jeffrey D. Bunce

Contribution from Tripos Associates, 1699 South Hanley Road, St. Louis, Missouri 63144. Received January 5, 1988

Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S.,

Generating Optimal Linear PLS Estimations (Golpe) - an Advanced Chemometric Tool for Handling 3D-QSAR Problems.

Quant Struct-Act Rel 1993, 12, (1), 9-20.

J. Med. Chem. 1994, 37, 2589-2601

2589

Comparative Molecular Field Analysis Using GRID Force-Field and GOLPE Variable Selection Methods in a Study of Inhibitors of Glycogen Phosphorylase δ

Gabriele Cruciani^{1*} and Kimberly A. Watson²

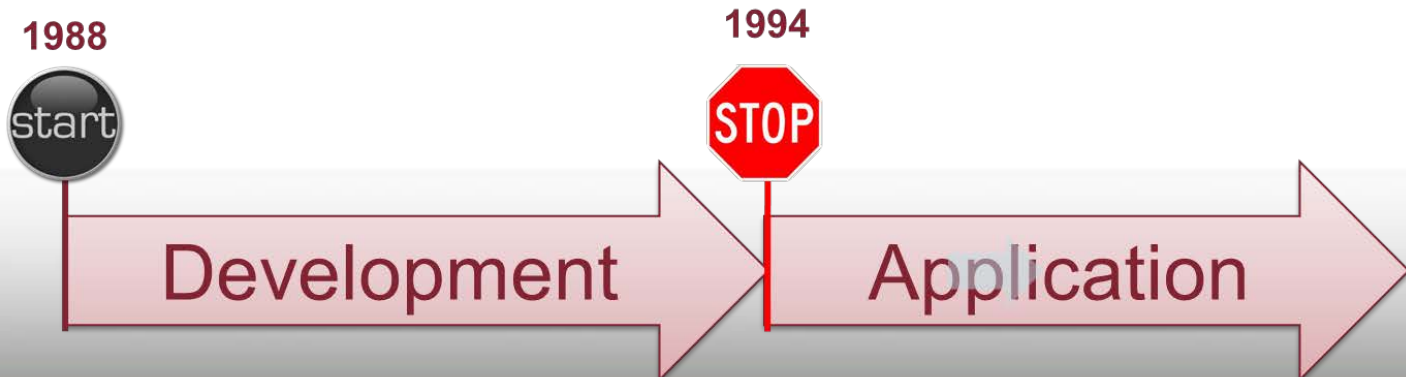
¹Department of Chemistry, University of Perugia, Via Elce di Sotto, 8, 06100 Perugia, Italy, and ²Laboratory of Molecular Biophysics, University of Oxford, South Parks Road, OX1 3QU, Oxford, England

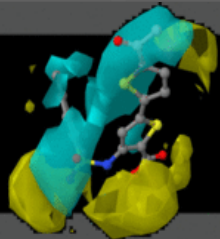
PCT		WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau	
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)			
(51) International Patent Classification ⁵ :	A1	(11) International Publication Number:	WO 92/22875
G06F 15/46		(43) International Publication Date:	23 December 1992 (23.12.92)
(21) International Application Number:	PCT/US91/04292	Published With international search report.	
(22) International Filing Date:	17 June 1991 (17.06.91)		
(71)(72) Applicants and Inventors: CRAMER, Richard, D., III [US/US]; Tripos Associates, Inc., 1699 S. Hanley Road, Ste. 303, St. Louis, MO 63144 (US). SVANTE, Wold [SE/US]; 371 Highland Avenue, Winchester, MA 01890 (US).			
(74) Agent: LIPTON, Robert, S.; Lipton & Famiglio, 201 N. Jackson Street, P.O. Box 546, Media, PA 19063 (US).			
(81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE, DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB, GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).			
(54) Title: COMPARATIVE MOLECULAR FIELD ANALYSIS (CoMFA)			



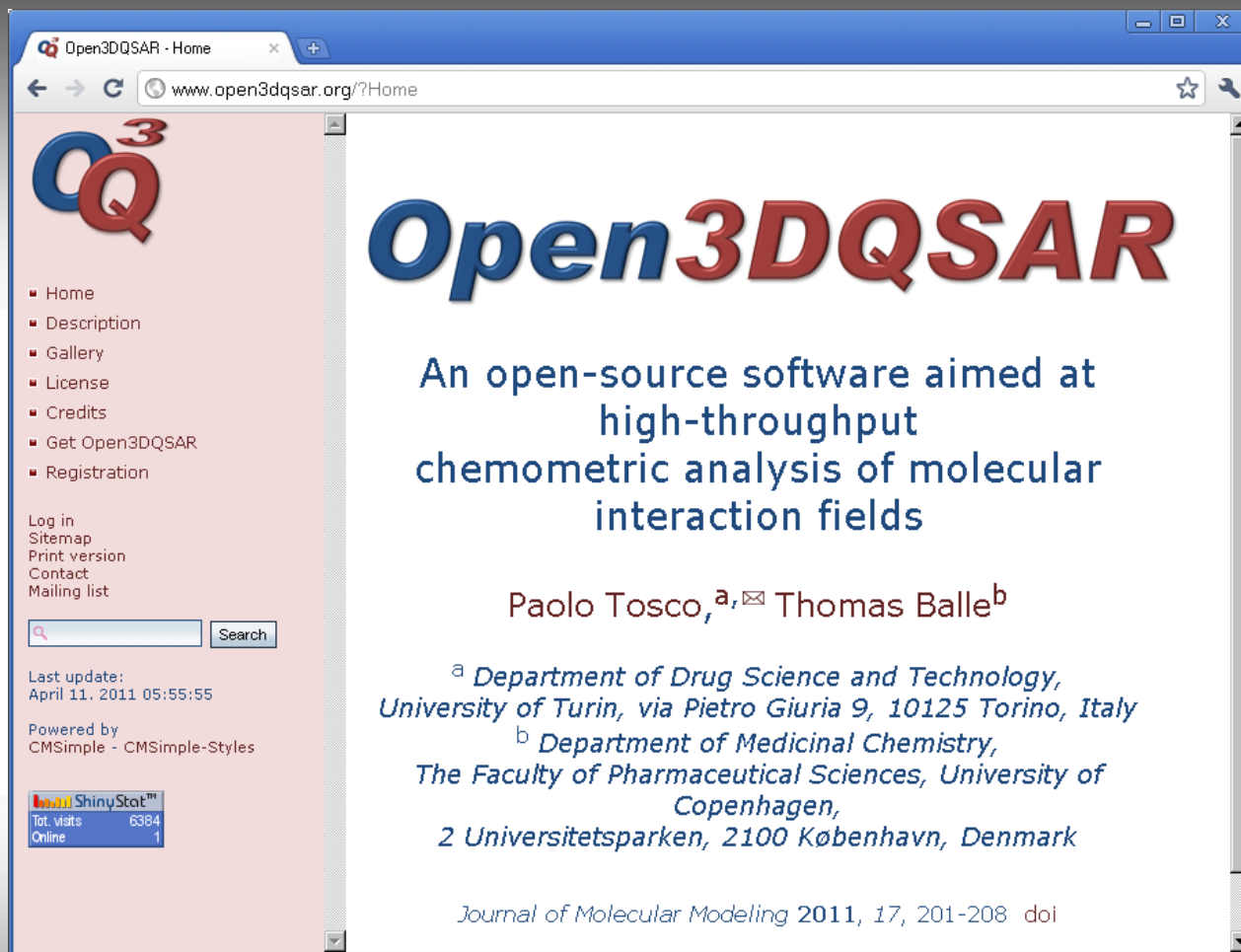
3-D QSAR Development

- 4-D QSAR, each molecule is represented by an ensemble of conformations, orientations, and protonation states
- 5-D QSAR, inclusion of the induced fit
- 6-D QSAR, the simultaneous evaluation of different solvation models
- CoMSIA
- VolSurf
-





2009: First Free 3D-QSARs



Open3DQSAR - Home

www.open3dqsar.org/?Home

Open3DQSAR

An open-source software aimed at high-throughput chemometric analysis of molecular interaction fields

Paolo Tosco,^{a,✉} Thomas Balle^b

^a Department of Drug Science and Technology,
University of Turin, via Pietro Giuria 9, 10125 Torino, Italy

^b Department of Medicinal Chemistry,
The Faculty of Pharmaceutical Sciences, University of
Copenhagen,
2 Universitetsparken, 2100 København, Denmark

Journal of Molecular Modeling 2011, 17, 201-208 doi

Navigation menu:

- Home
- Description
- Gallery
- License
- Credits
- Get Open3DQSAR
- Registration

Log in
Sitemap
Print version
Contact
Mailing list

Search

Last update:
April 11, 2011 05:55:55

Powered by
CMSimple - CMSimple-Styles

ShinyStat™	
Tot. visits	6384
Online	1



2009: Free Software/Open Source Based 3D-QSARs

rcmd
www.rcmd.it

www.rcmd.it
rcmd
Rome Center For MOLECULAR DESIGN

SAPIENZA
UNIVERSITÀ DI ROMA
Facoltà di Farmacia

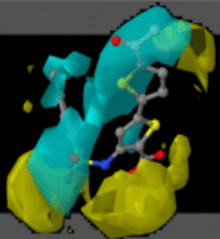
VEGFR-2 Inhibitors. Ligand-Based, Structure-Based and 3-D QSAR studies as tools to Design New Small Molecules

Rino Ragno
Rome Center for Molecular Design, Dipartimento di Chimica e Tecnologia del Farmaco, Università degli Studi di Roma "La Sapienza", P.le A. Moro 5, 00185 Roma, Italy

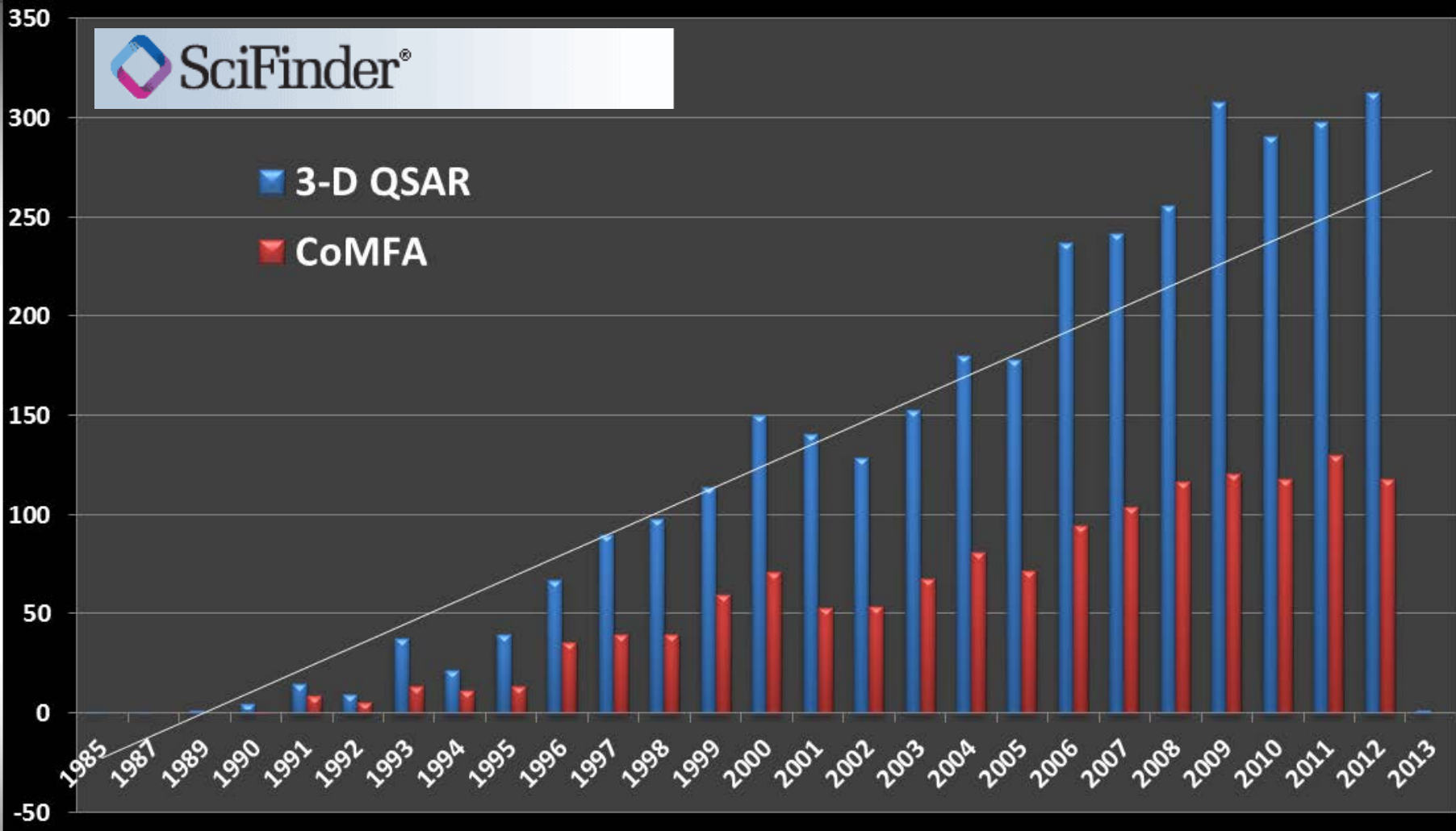
Manuscript Submitted

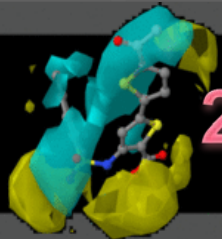


XXXVII Congresso Nazionale di Chimica Inorganica
XXIII Congresso Nazionale della Società Chimica
Italiana,
Sorrento, 5 - 10 Luglio 2009.



3D-QSAR is Alive!





2011: The First 3D-QSAR Server



On 17 June 2011 the patent PLS+MIF restriction dropped and now a new 3-D QSAR explosion is expected.

In view of this event, five years ago, at RCMD, we started an ambitious project aimed to build a 3-D QSAR web server using just open source or free software

www.3D-QSAR.com
THREE-DIMENSIONAL QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIPS SERVER



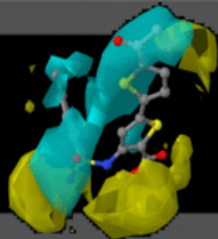


The Server Engine



3-D QSAutogrid/R

Only Open Source Software!




The Server Engine

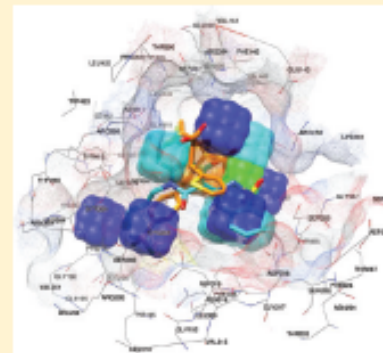
3-D QSAutogrid/R: An Alternative Procedure To Build 3-D QSAR Models. Methodologies and Applications

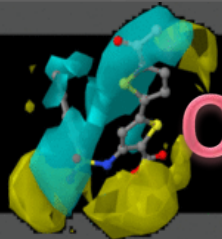
Flavio Ballante[†] and Rino Ragno^{*†}

[†]Rome Center for Molecular Design, Dipartimento di Chimica e Tecnologie del Farmaco, Sapienza Università di Roma, P. le A. Moro 5, 00185, Rome, Italy

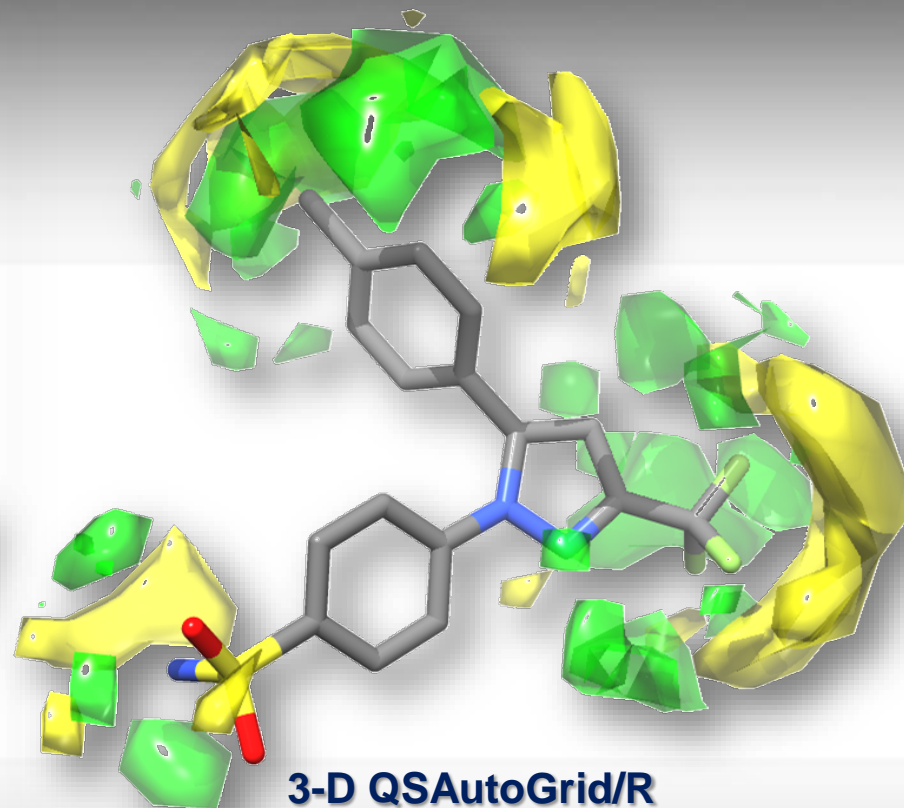
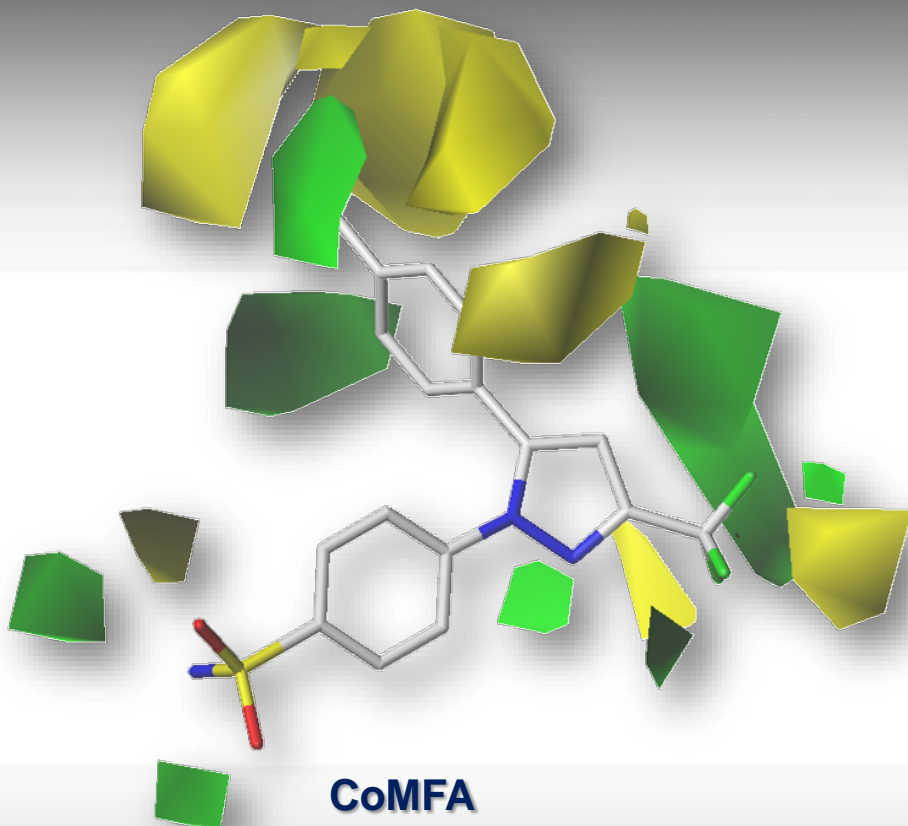
 Supporting Information

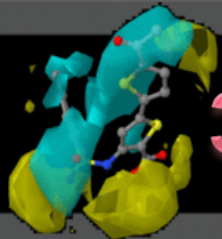
ABSTRACT: Since it first appeared in 1988 3-D QSAR has proved its potential in the field of drug design and activity prediction. Although thousands of citations now exist in 3-D QSAR, its development was rather slow with the majority of new 3-D QSAR applications just extensions of CoMFA. An alternative way to build 3-D QSAR models, based on an evolution of software, has been named 3-D QSAutogrid/R and has been developed to use only software freely available to academics. 3-D QSAutogrid/R covers all the main features of CoMFA and GRID/GOLPE with implementation by multiprobe/multiregion variable selection (MPGRS) that improves the simplification of interpretation of the 3-D QSAR map. The methodology is based on the integration of the molecular interaction fields as calculated by AutoGrid and the R statistical environment that can be easily coupled with many free graphical molecular interfaces such as UCSF-Chimera, AutoDock Tools, Jmol, and others. The description of each R package is reported in detail, and, to assess its validity, 3-D QSAutogrid/R has been applied to three molecular data sets of which either CoMFA or GRID/GOLPE models were reported in order to compare the results. 3-D QSAutogrid/R has been used as the core engine to prepare more than 240 3-D QSAR models forming the very first 3-D QSAR server (www.3d-qsar.com) with its code freely available through R-Cran distribution.



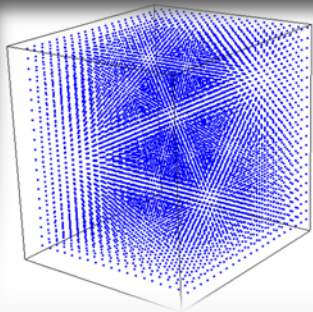


CoMFA vs 3-D QSAutoGrid/R

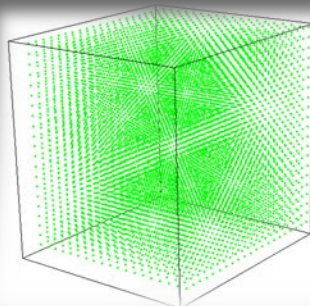




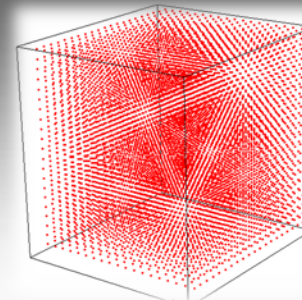
3-D QSAutoGrid/R → MPGRS



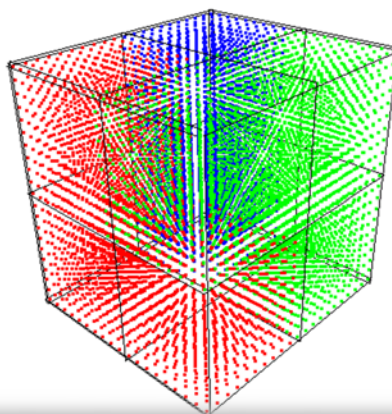
GRS

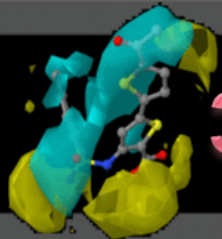


GRS

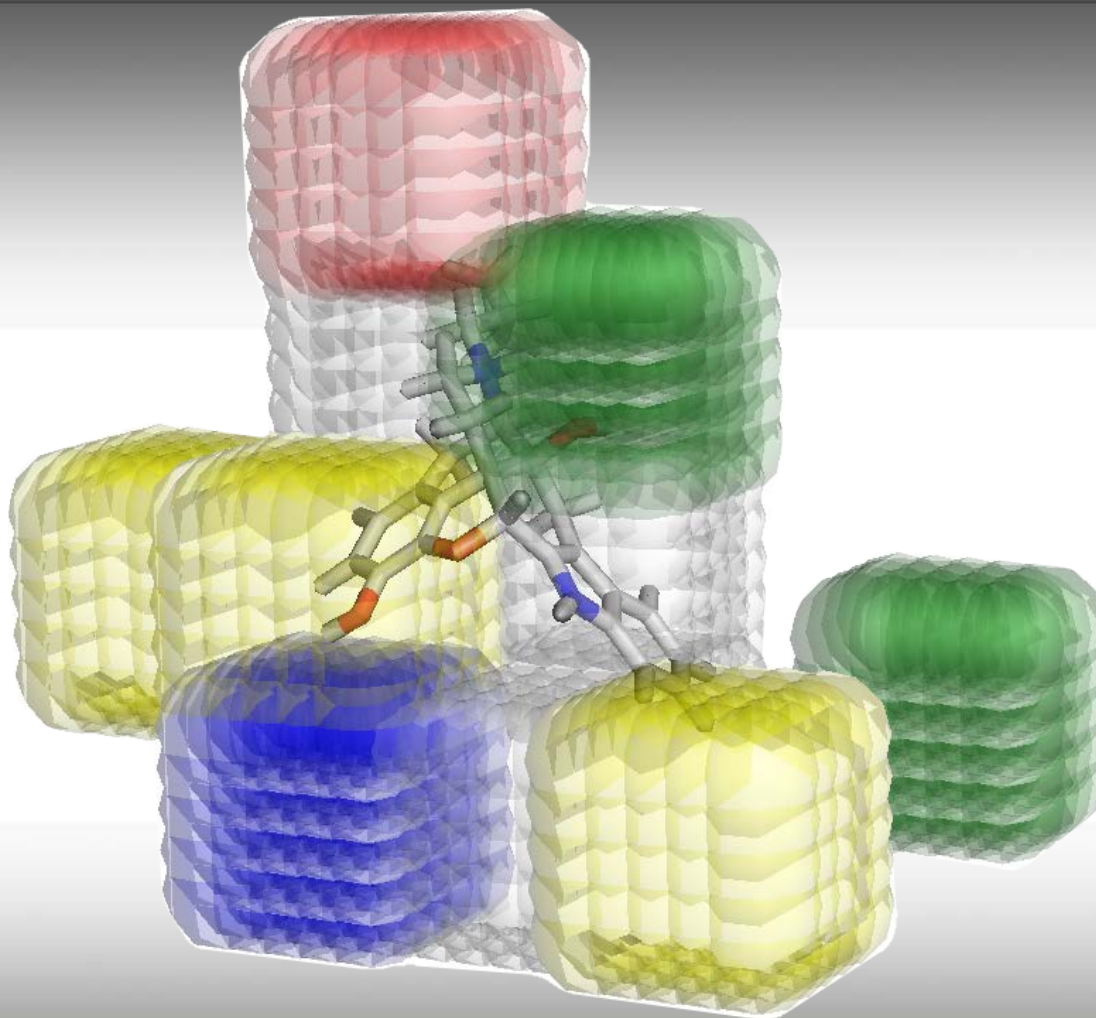


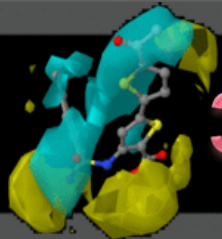
GRS



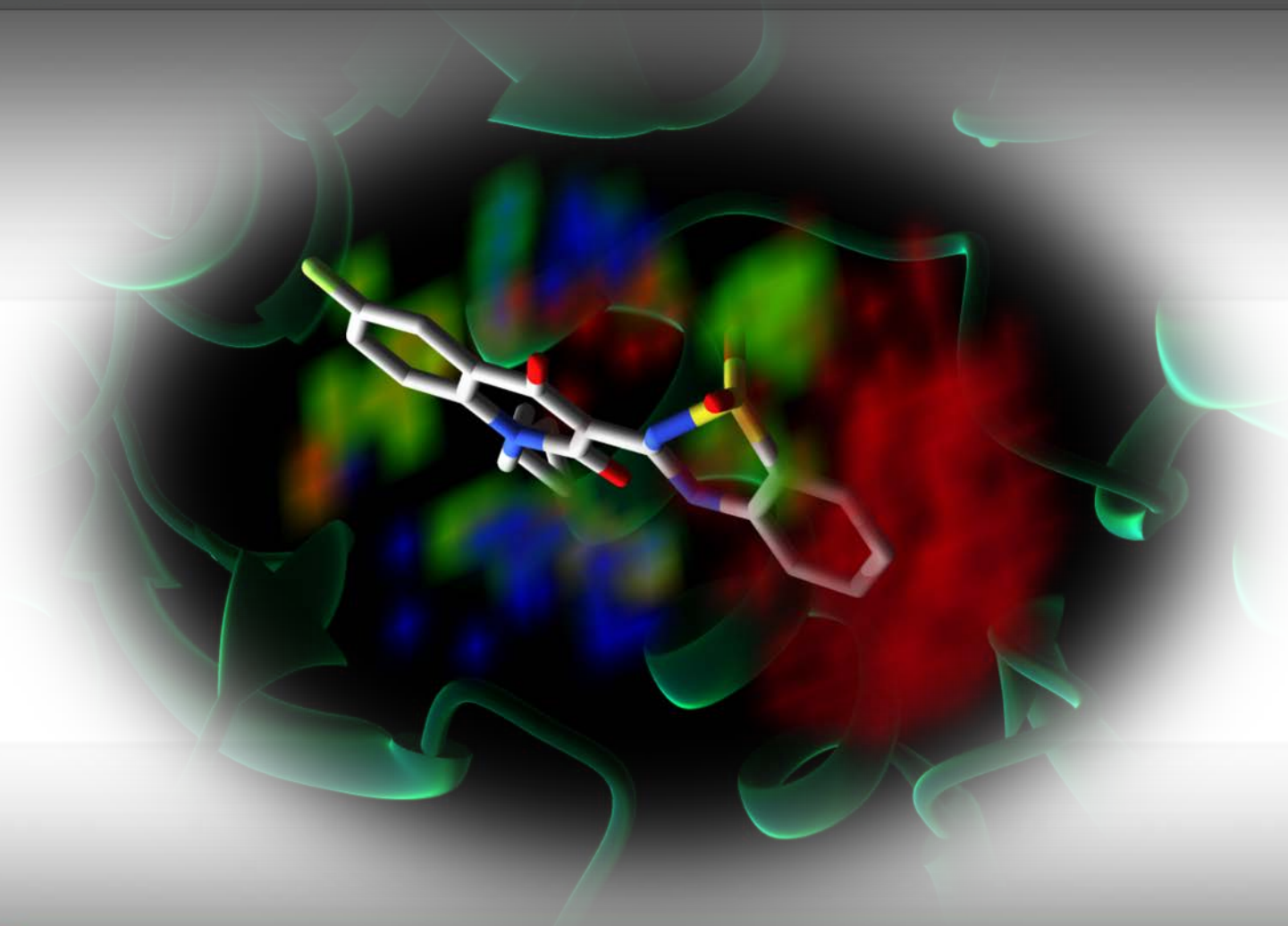


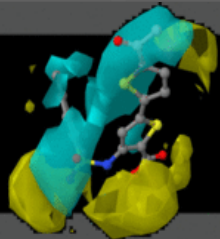
3-D QSAutoGrid/R → MPGRS





3-D QSAutoGrid/R → MPGRS





Live View

www.3d-qsar.com

3D-QSAR
THREE-DIMENSIONAL QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIPS SERVER

SAPIENZA
UNIVERSITÀ DI ROMA
by **RCMD**

Home Launch Predictor **Datasets** About Change Log

You are inspecting model AchE_A

Model Description

The AchE dataset was taken from J. Med. Chem. 2004, 47, 5541-5554. Paper abstract: A large number of methods are available for modeling quantitative structure-activity relationships (QSAR). We examine the predictive accuracy of several methods applied to data sets of inhibitors for angiotensin converting enzyme, acetylcholinesterase, benzodiazepine receptor, cyclooxygenase-2, dihydrofolate reductase, glycogen phosphorylase b, thermolysin, and thrombin. Descriptors calculated with CoMFA, CoMSIA, EVA, HQSAR, and traditional 2D and 2.5D descriptors were used for developing models with partial least squares (PLS). In addition, the genetic function approximation algorithm, genetic PLS, and back-propagation neural networks were used for deriving models from 2.5D descriptors (i.e., 2D descriptors and 3D descriptors calculated from CORINA structures and Gastelger-Marsili charges). Predictive accuracy was assessed using designed test sets. It was found that HQSAR generally performs as well as CoMFA and CoMSIA; other descriptor sets performed less well. When 2.5D descriptors were used, only neural network ensembles were found to be similarly or more predictive than PLS models. In addition, we show that many cross-validation procedures yield similar estimates of the interpolative accuracy of methods. However, the lack of correspondence between cross-validated and test set predictive accuracy for four sets underscores the benefit of using designed test sets.

AchE Models

- AchE_e
- AchE_d
- AchE_C
- AchE_A
- AchE_HQ
- AchE_N

Other Datasets

- AG
- ACE
- AchE
- Al
- ARB
- ATA

Training Set

Show	Mol id	Exp p[act]
<input type="checkbox"/>	94	9.520
<input type="checkbox"/>	93	9.220
<input type="checkbox"/>	141	8.920
<input type="checkbox"/>	128	8.700
<input type="checkbox"/>	140	8.550
<input type="checkbox"/>	77	8.380
<input type="checkbox"/>	81	8.270
<input type="checkbox"/>	117	8.240
<input type="checkbox"/>	118	8.190