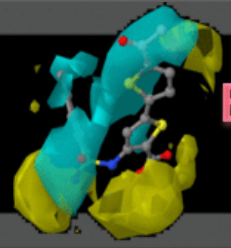
# Ligand-Based Drug Design


SAPIENZA
UNIVERSITÀ DI ROMA



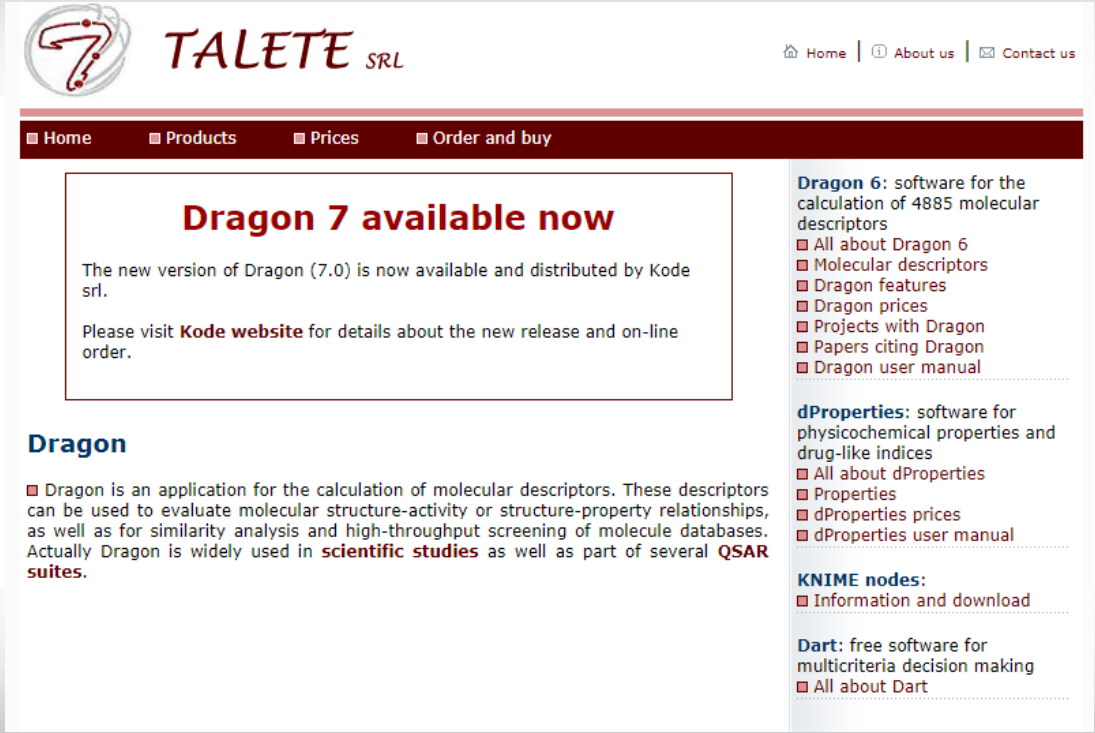
# EVALUATION AND VALIDATION OF A MODEL

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



# 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



*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.

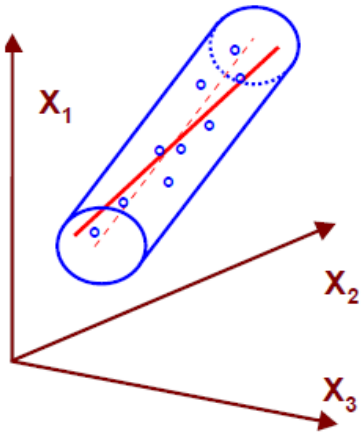
**Description of Molecules with Molecular Interaction Fields (MIF)**

# PCA

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

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

### PLS (Partial Least Squares) Analysis

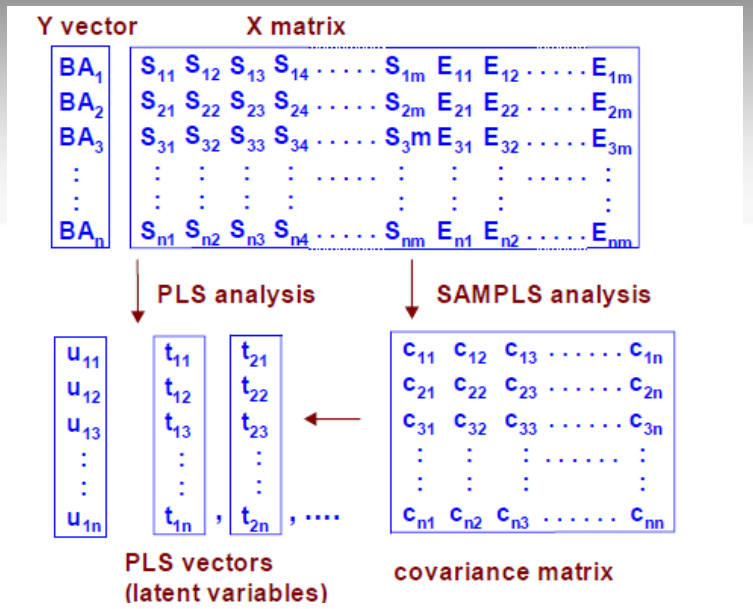



From u = kt

(k = constants $k_1$, $k_2$ ... $k_j$;
j = number of PLS vektors)

follows:

$$BA_i = a_1S_{i1} + a_2S_{i2} + a_3S_{i3} + \ldots + a_mS_{im} + b_1E_{i1} + b_2E_{i2} + b_3E_{i3} + \ldots + b_mE_{im}$$

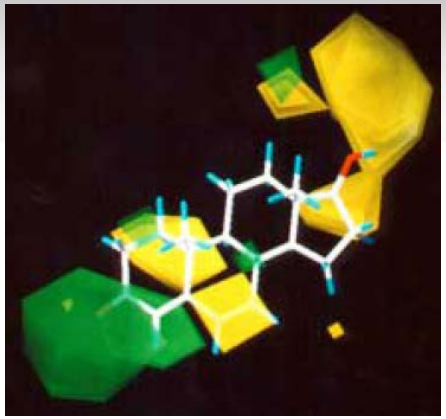


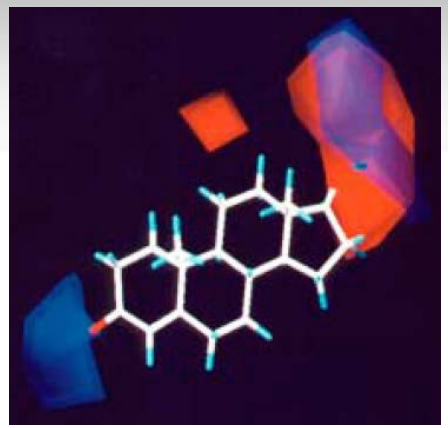

**Reduction of Dimensionality into Few New Highly Informative Entities**

# ----- Principal Components -----

# 1988: The First 3-D QSAR

## PLS + MIF → CoMFA!





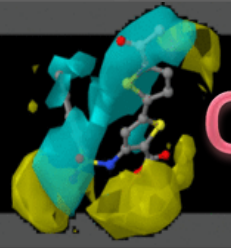
*J. Am. Chem. Soc.* **1988**, *110*, 5959–5967



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*

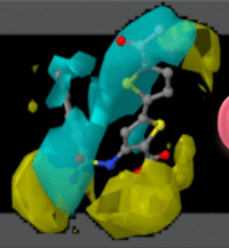


# CoMFA/3-D QSAR Procedure

- 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

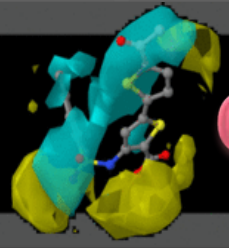
Training Set

Molecular Alignment

Molecular Interaction Fields (MIF)

Model Generation

Internal and External Validation



# CoMFA/3-D QSAR Procedure

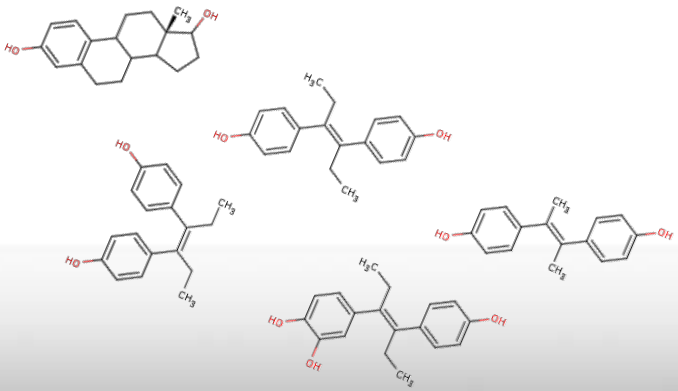
**Training Set**

Molecular lignment

Molecular Interaction Fields (MIF)

Model Generation

Internal and External Validation



# CoMFA/3-D QSAR Procedure

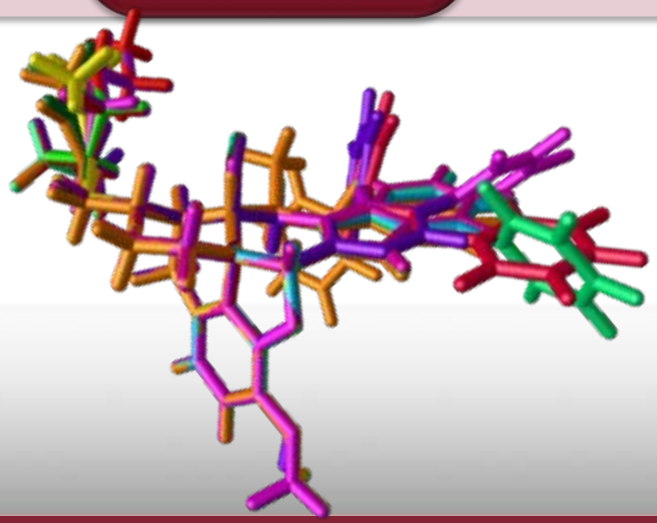
Training Set

**Molecular Alignment**

Molecular Interaction Fields (MIF)

Model Generation

Internal and External Validation



# CoMFA/3-D QSAR Procedure

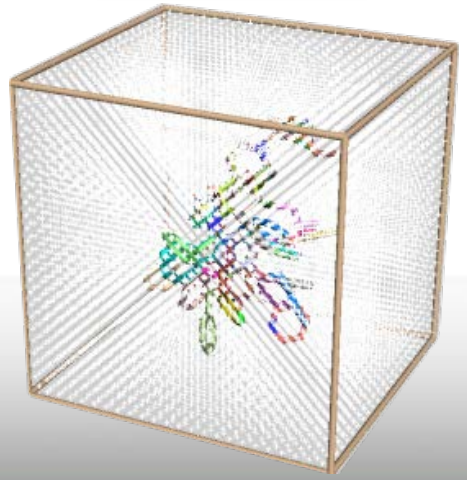
Training Set

Molecular Alignment

**Molecular Interaction Fields (MIF)**

Model Generation

Internal and External Validation



# CoMFA/3-D QSAR Procedure

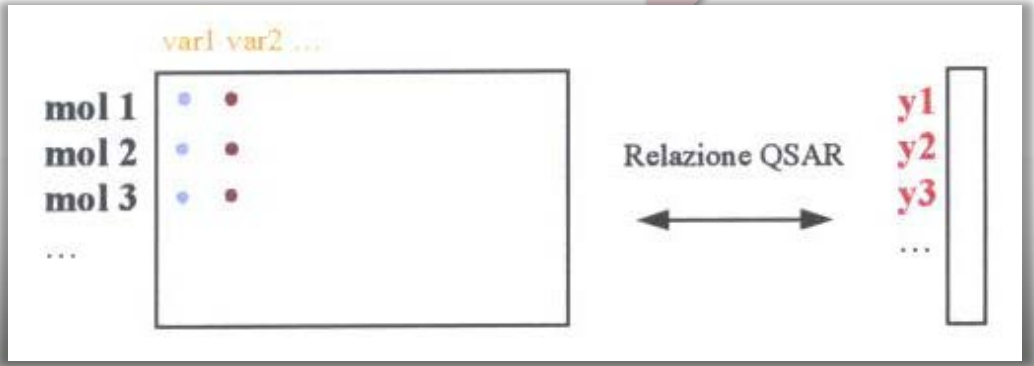
Training Set

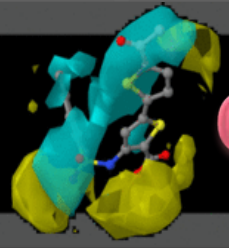
Molecular lignment

Molecular Interaction Fields (MIF)

**Model Generation**

Internal and External Validation





# CoMFA/3-D QSAR Procedure

Training Set

Molecular Alignment

Molecular Interaction Fields (MIF)

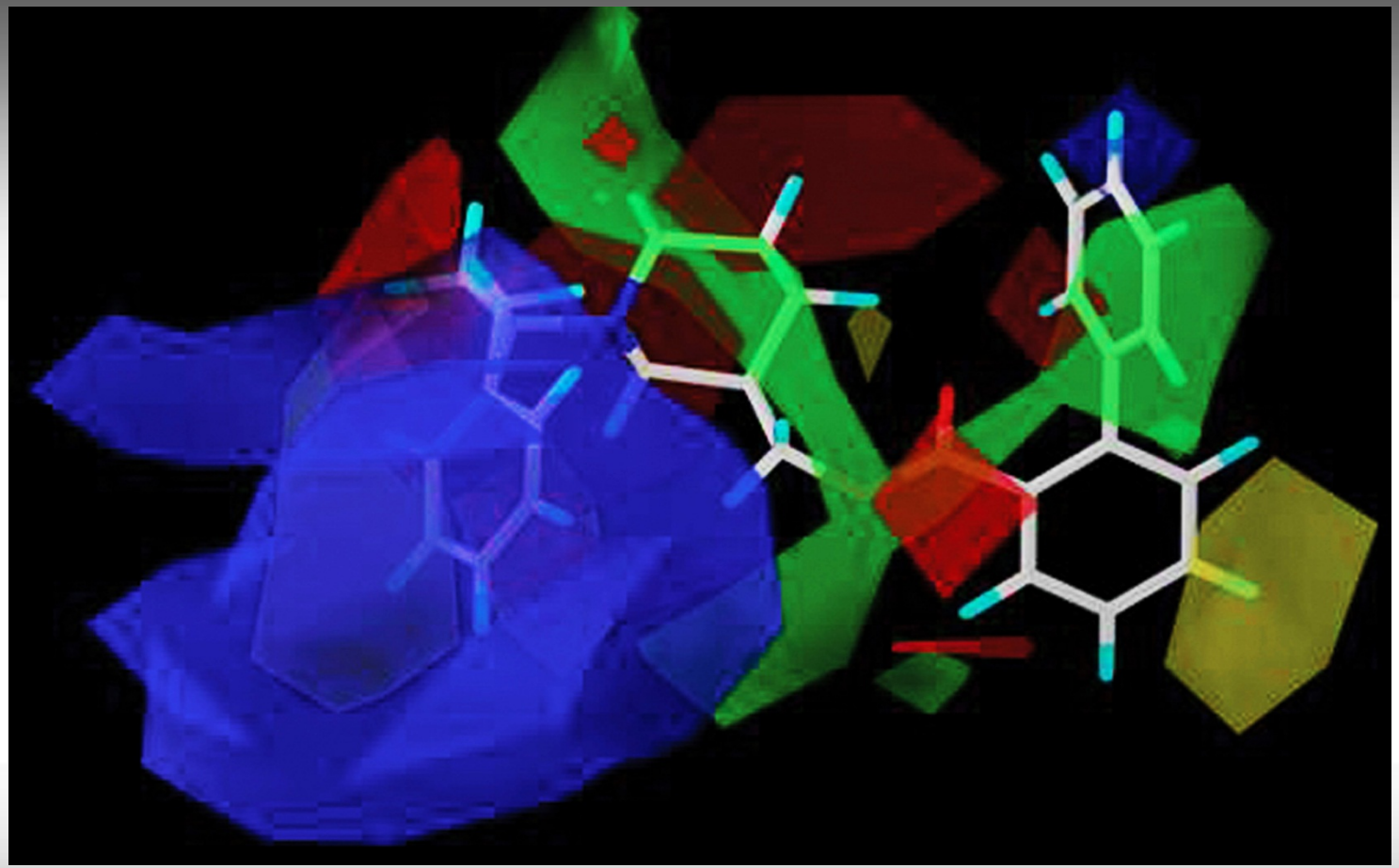
Model Generation

**Internal and External Validation**

$$SDEP = \sqrt{\sum \left(y_{\exp} - y_{pred}\right)^2 / n - 1}$$

$$q^2 = 1 - \sum \left(y_{\exp} - y_{pred}\right)^2 / \sum \left(y_{\exp} - y\right)^2$$

# CoMFA/3-D QSAR Procedure



# CoMFA Limits

## CRUCIAL POINTS!

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

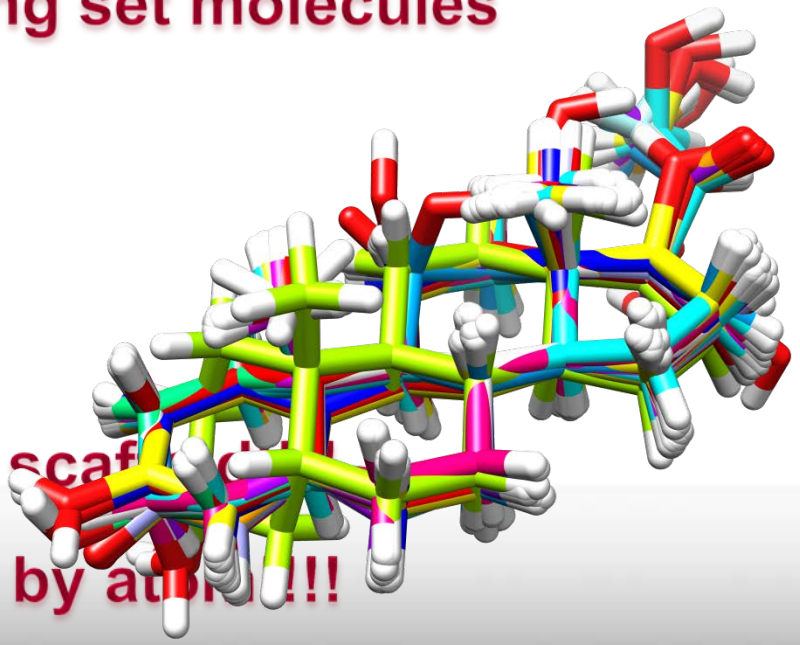
## Original CoMFA Tricks:

- 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 scaf[illegible]
- Directed superimposed atom by at[illegible]!!!

# 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 *b***

Gabriele Cruciani*,† and Kimberly A. Watson‡

*Department of Chemistry, University of Perugia, Via Elce di Sotto, 8, 06100 Perugia, Italy, and Laboratory 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 :** G06F 15/46 | A1 | **(11) International Publication Number:** WO 92/22875<br>**(43) International Publication Date:** 23 December 1992 (23.12.92) |

**(21) International Application Number:** PCT/US91/04292
**(22) International Filing Date:** 17 June 1991 (17.06.91)

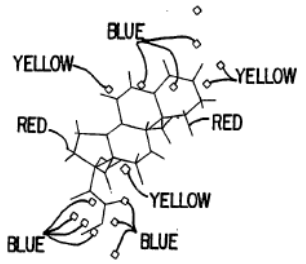
**Published**
*With international search report.*

**(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).

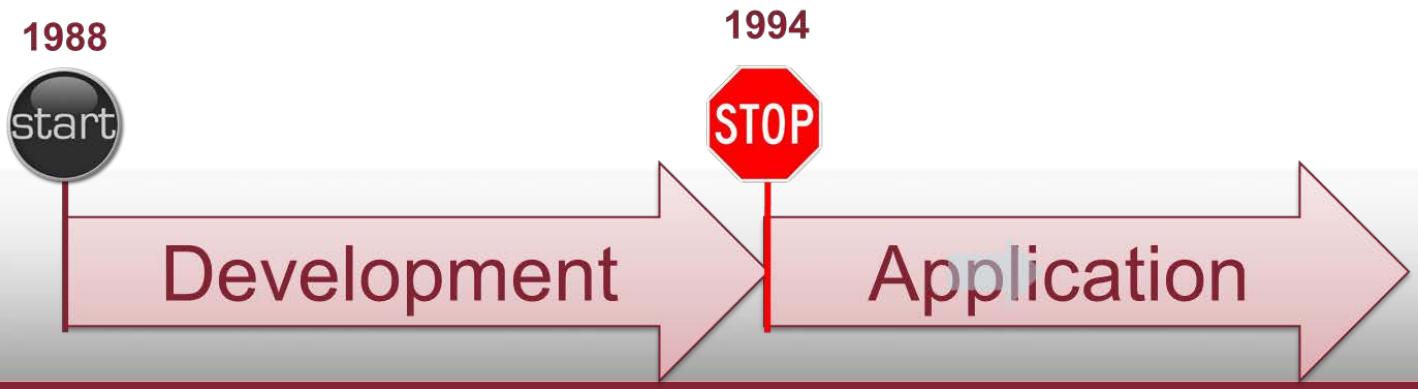
**(54) Title:** COMPARATIVE MOLECULAR FIELD ANALYSIS (CoMFA)



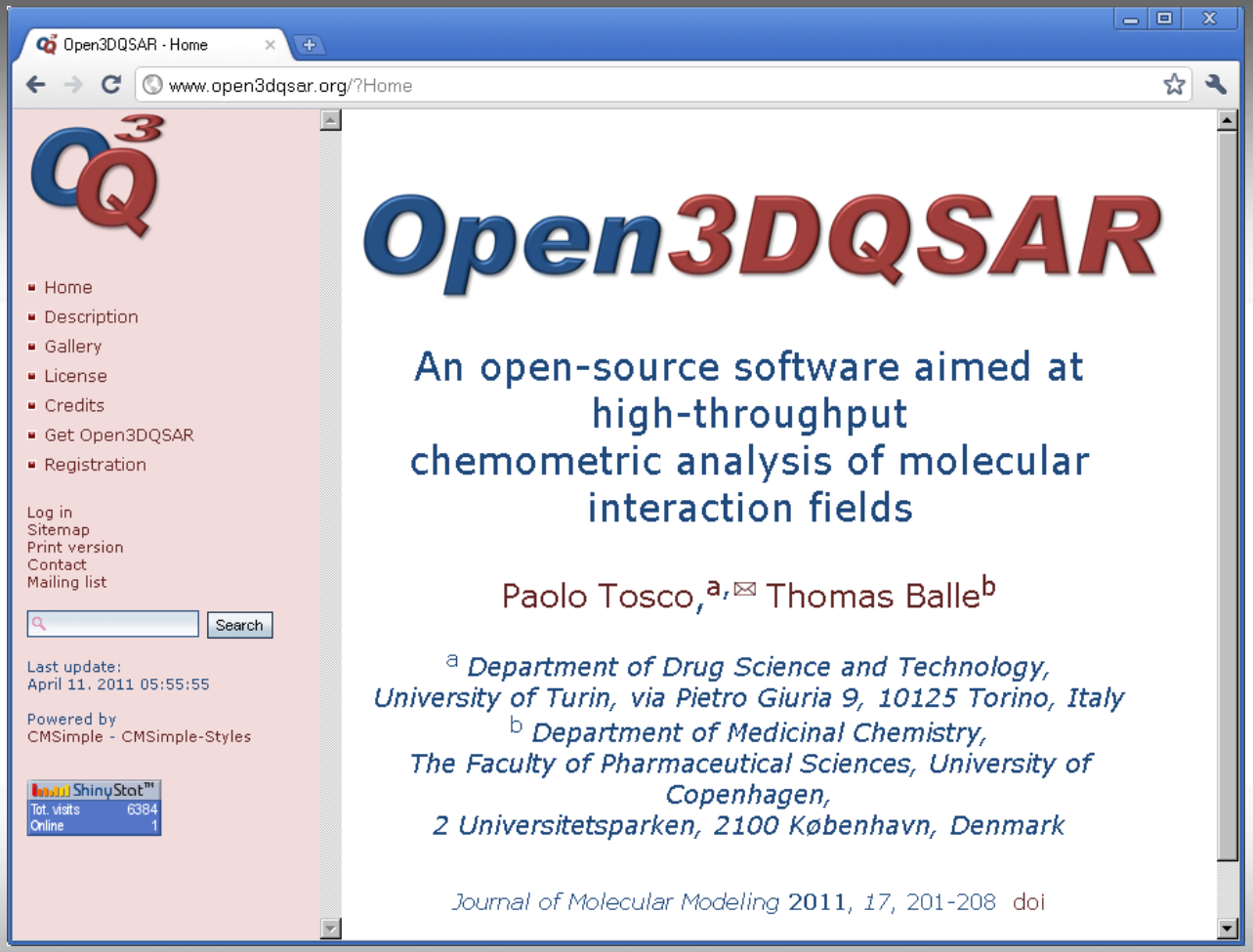
# 3-D QSAR Development

- 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

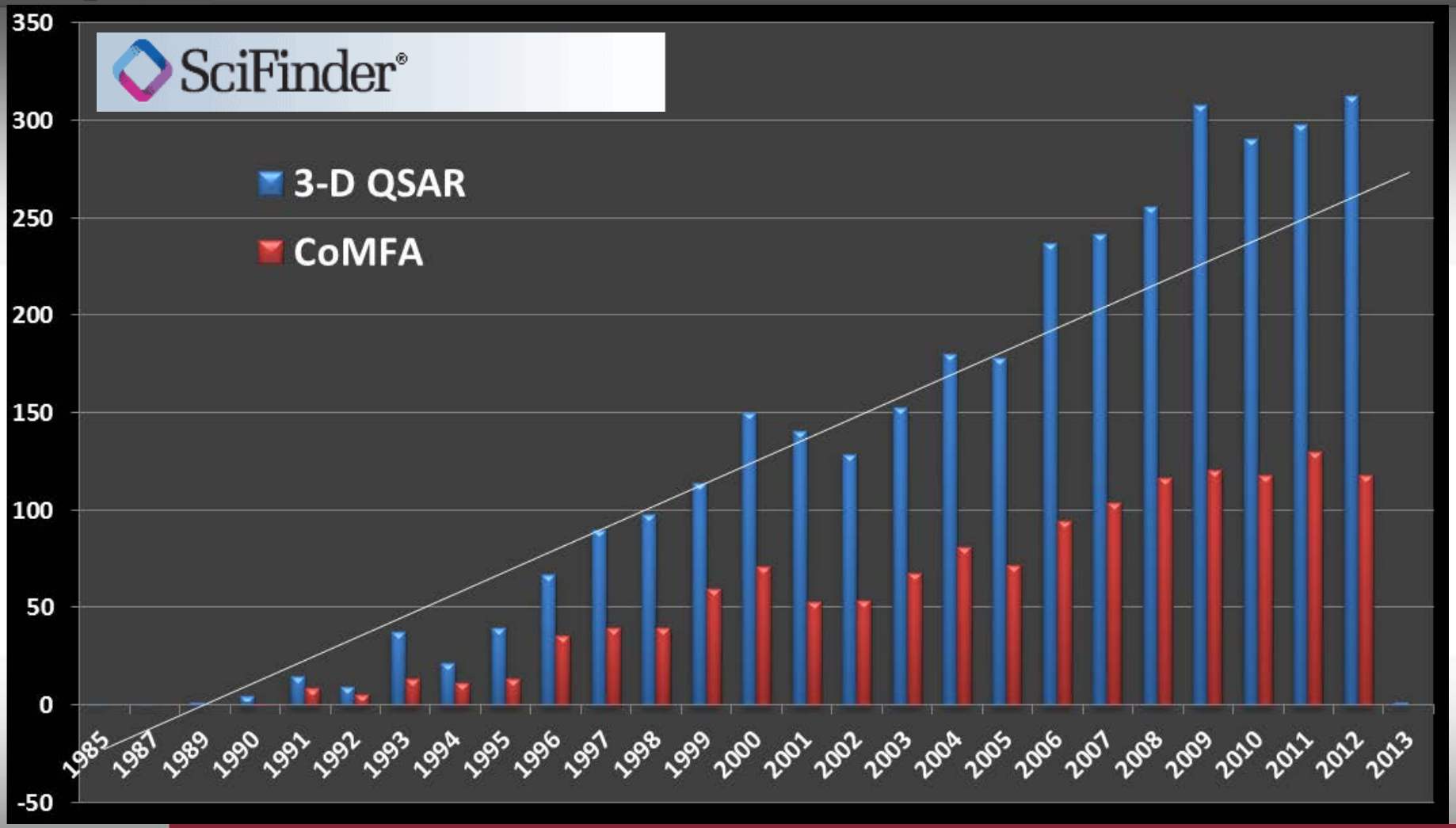


# 2009: Free Software/Open Source Based 3D-QSARs




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

# 3D-QSAR is Alive!

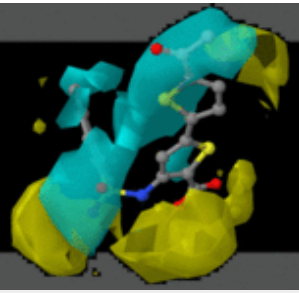


# 2011: The First 3D-QSAR Server

**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



Sapienza Università di Roma

# The Server Engine

## 3-D QSAutogrid/R

## Only Open Source Software!

# The Server Engine

JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

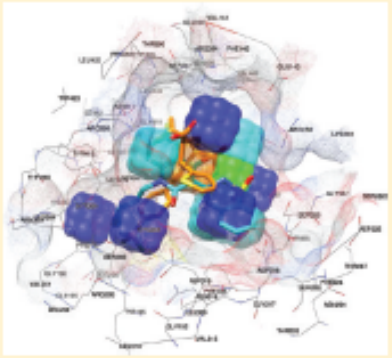
## 3-D QSAutogrid/R: An Alternative Procedure To Build 3-D QSAR Models. Methodologies and Applications

Flavio Ballante[†] and Rino Ragno*[,†]

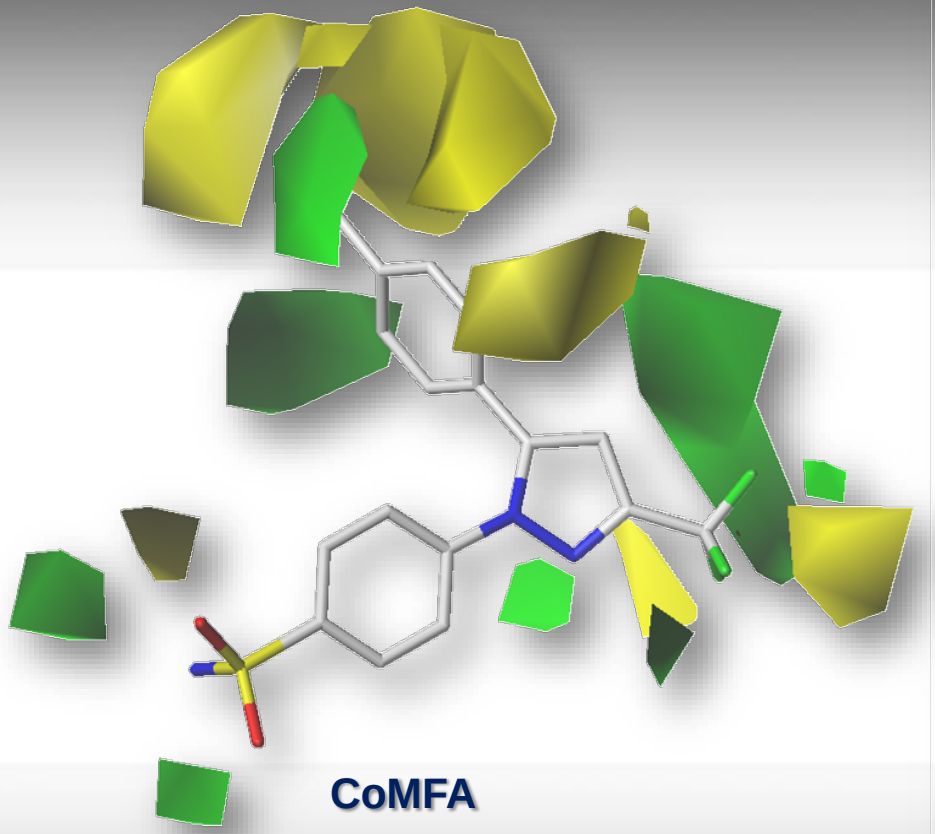
[†]Rome Center for Molecular Design, Dipartimento di Chimica e Tecnologie del Farmaco, Sapienza Università di Roma, P. le A. Moro 5, 00185, Rome, Italy

Supporting Information

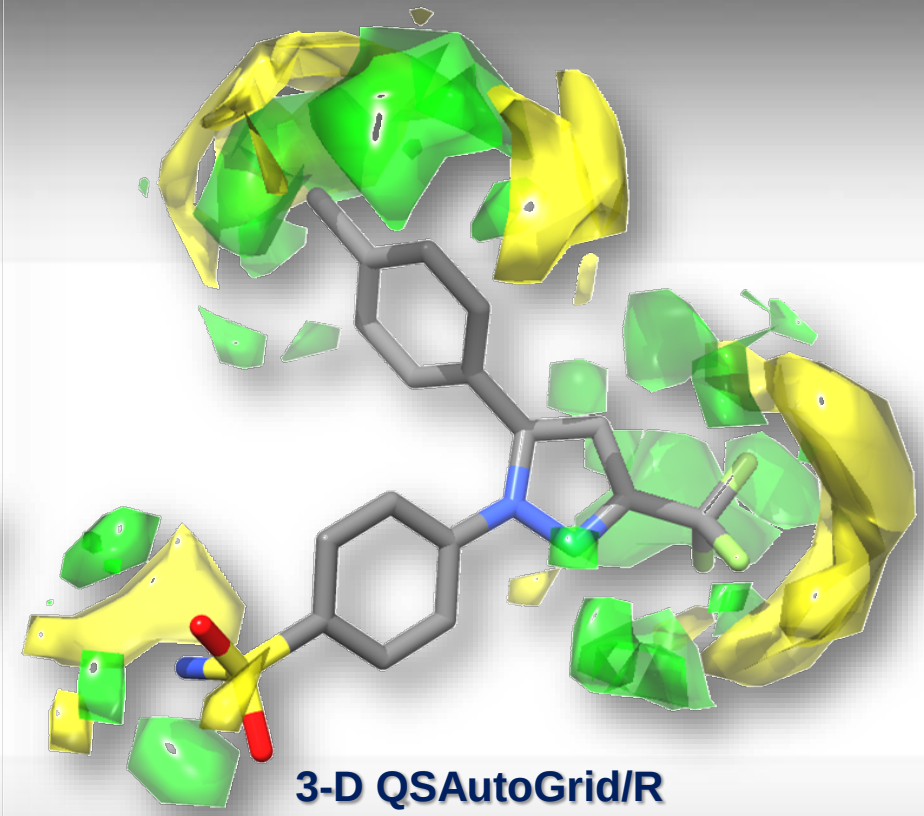
**ABSTRACT:** Since it first appeared in 1988 3-D QSAR has proved its potential in the field of drug design and activity prediction. Although thousands of citations now exist in 3-D QSAR, its development was rather slow with the majority of new 3-D QSAR applications just extensions of CoMFA. An alternative 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 that 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

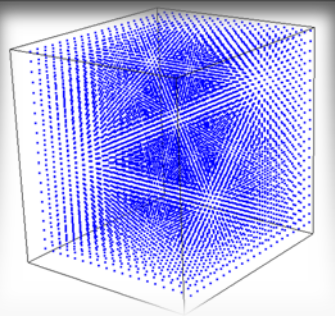


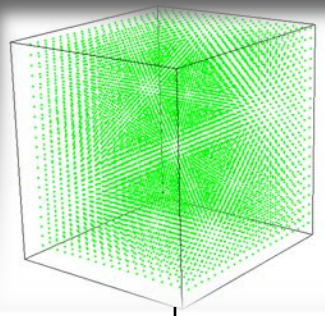
CoMFA

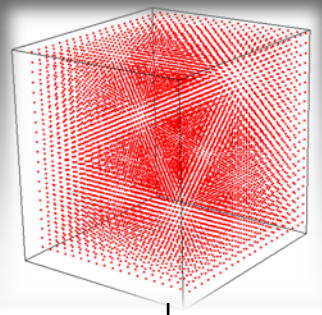


3-D QSAutoGrid/R

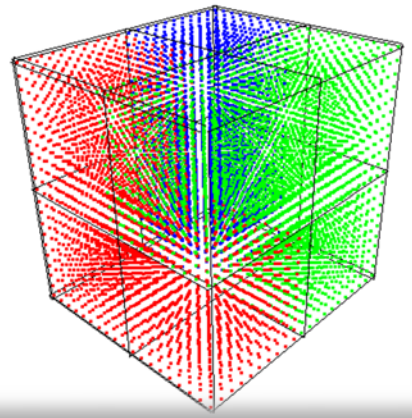
# 3-D QSAutoGrid/R → MPGRS



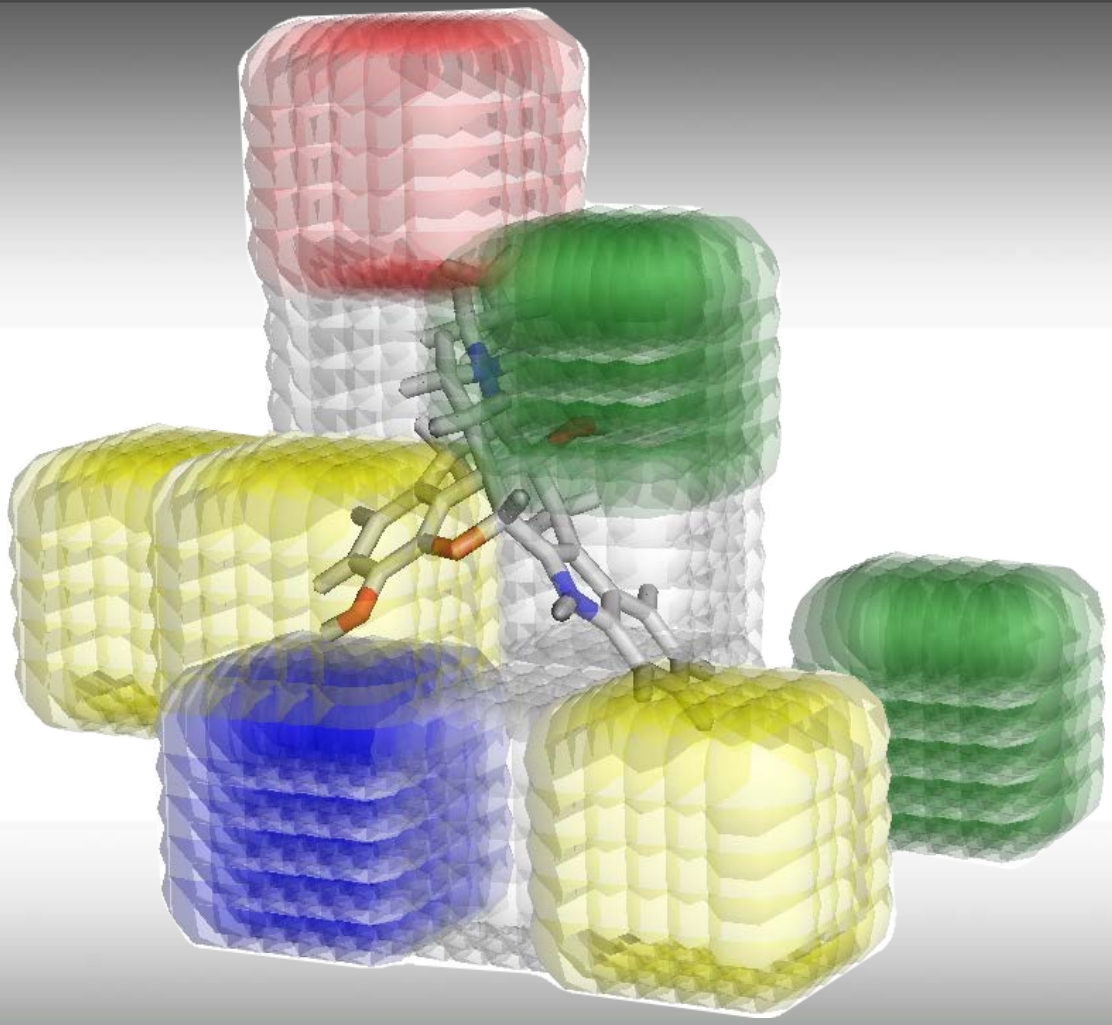




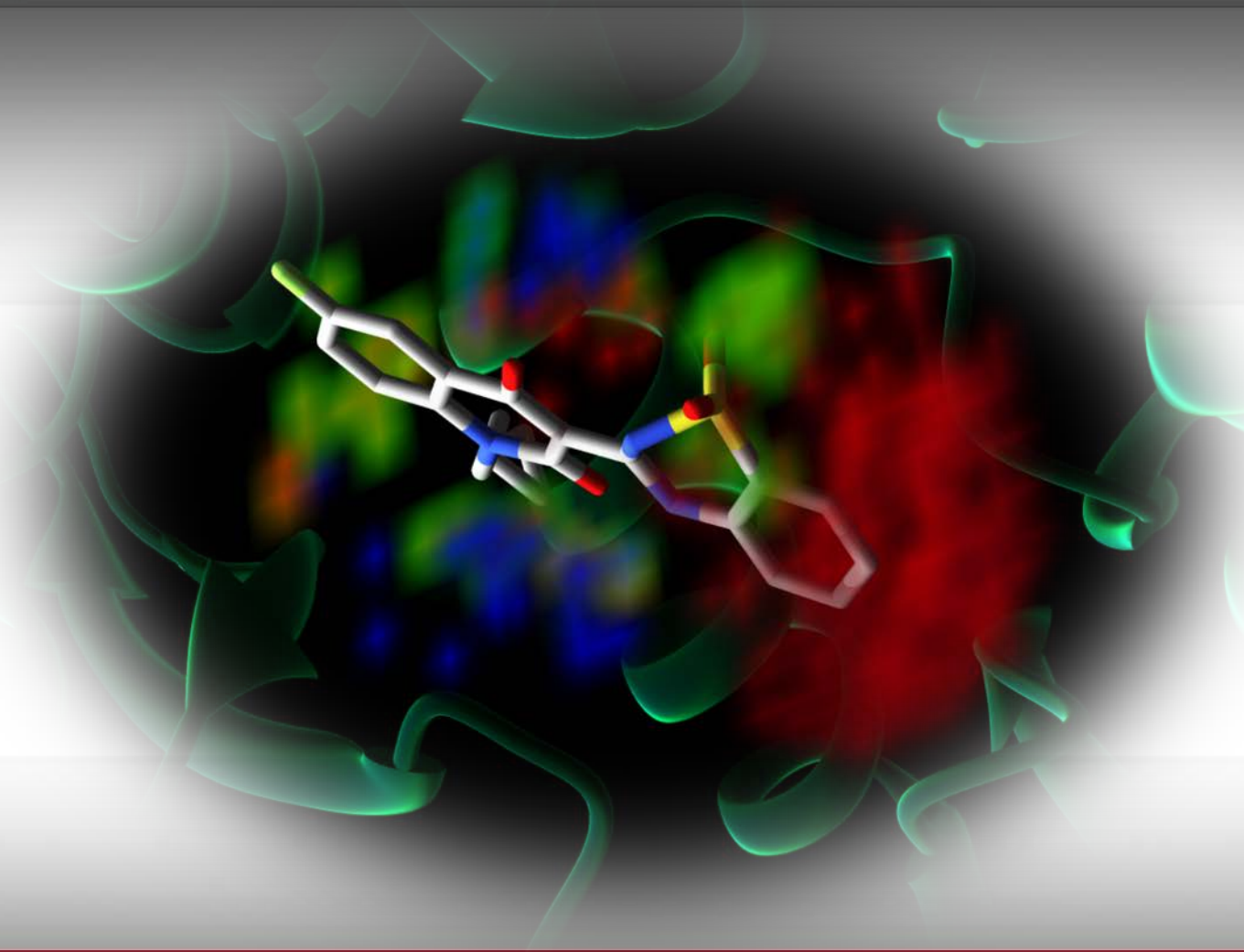
GRS GRS GRS



# 3-D QSAutoGrid/R → MPGRS



# 3-D QSAutoGrid/R → MPGRS



# Live View

## www.3d-qsar.com