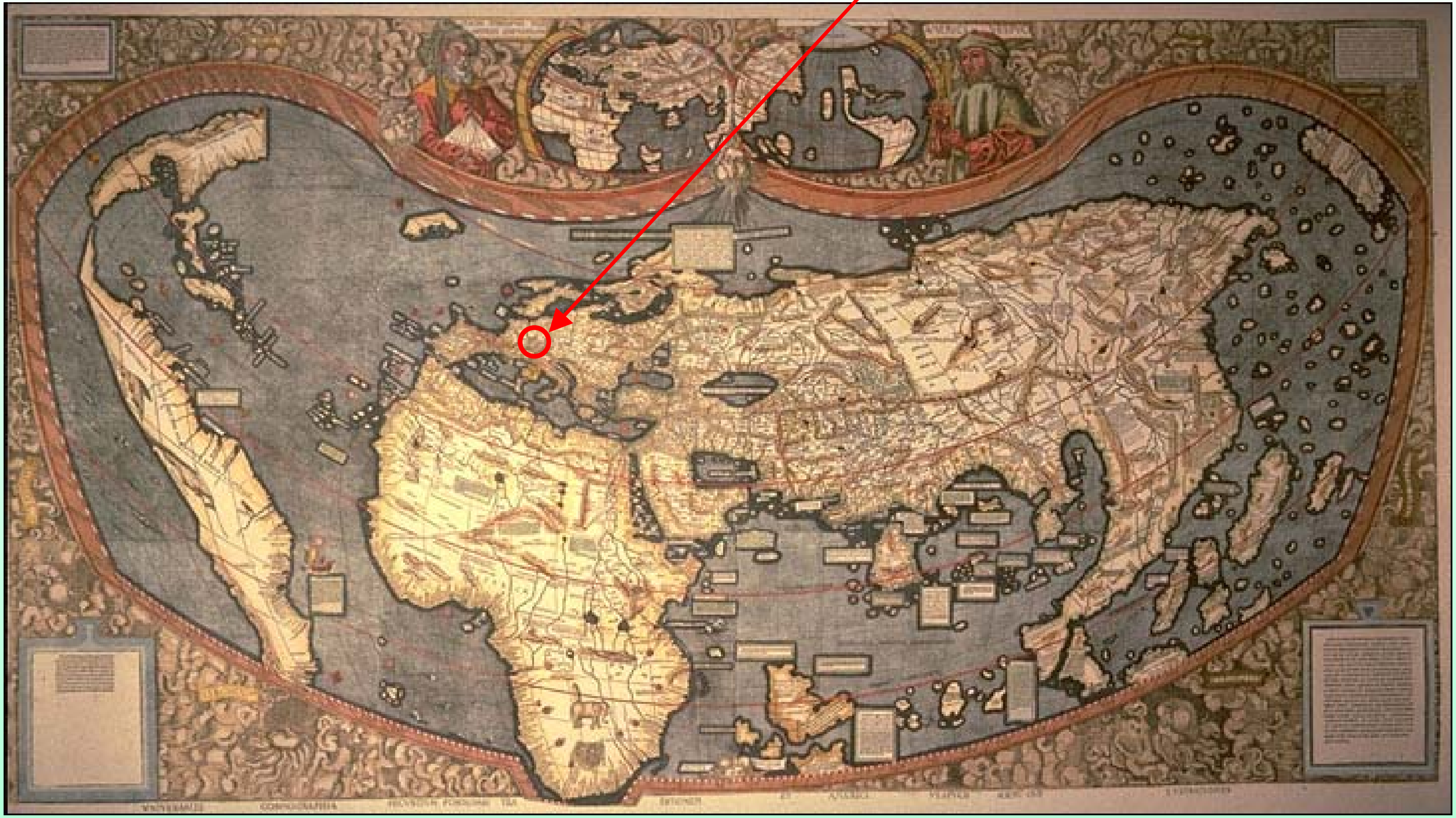


Obernai

Martin Waldseemüller's World Map of 1507; the FIRST map to use the name "America" to label the New World

Molecular descriptors

An introduction

Prof. Roberto Todeschini

Dr. Davide Ballabio

Dr. Viviana Consonni

Dr. Alberto Manganaro

Dr. Andrea Mauri

The chemical data

- synthesis: chemistry produces the objetcs of its own study

- chemical composition: a unifying concept for all the experimental sciences

- molecular structure: one the most fruitful scientific concepts of this century

Molecular structure

The concept of molecular structure is one of the most reach of this century.

Molecular structure

The basic assumptions are that different molecular structures have different chemical properties and similar molecular structures have similar molecular properties.

Molecular structure

Each molecular representation represents a different way to look at the molecular structure and its chemical meaning is strongly immersed in the framework of the chemical theories.

Some historical notes

“... : benchè certamente si traveggano già dei rapporti fra la costituzione chimica (composizione e struttura) e le proprietà fisiche loro, è ancor certamente di gran lunga troppo ristretto il numero dei fatti, per dedurne delle conseguenze, che oltre al carattere d’una semplice ipotesi possono pretendere anche quello della probabilità.

In ogni caso tali rapporti non sono di natura tanto semplice come a priori forse era lecito aspettarsi.

Di certo le proprietà fisiche dei corpi sono in primo luogo una funzione della composizione e struttura loro, sulla di cui forma nulla ancora si sa; funzione probabilmente molto complessa e per il di cui studio occorrerà un imprevedibile numero di fatti, onde poter sufficientemente restringere la cerchia delle rappresentazioni possibili.”

Some historical notes

Some historical notes

Studi sull’isomeria delle così dette sostanze aromatiche a sei atomi di carbonio.

Gazzetta Chimica Italiana, vol. IV, p.305

1874

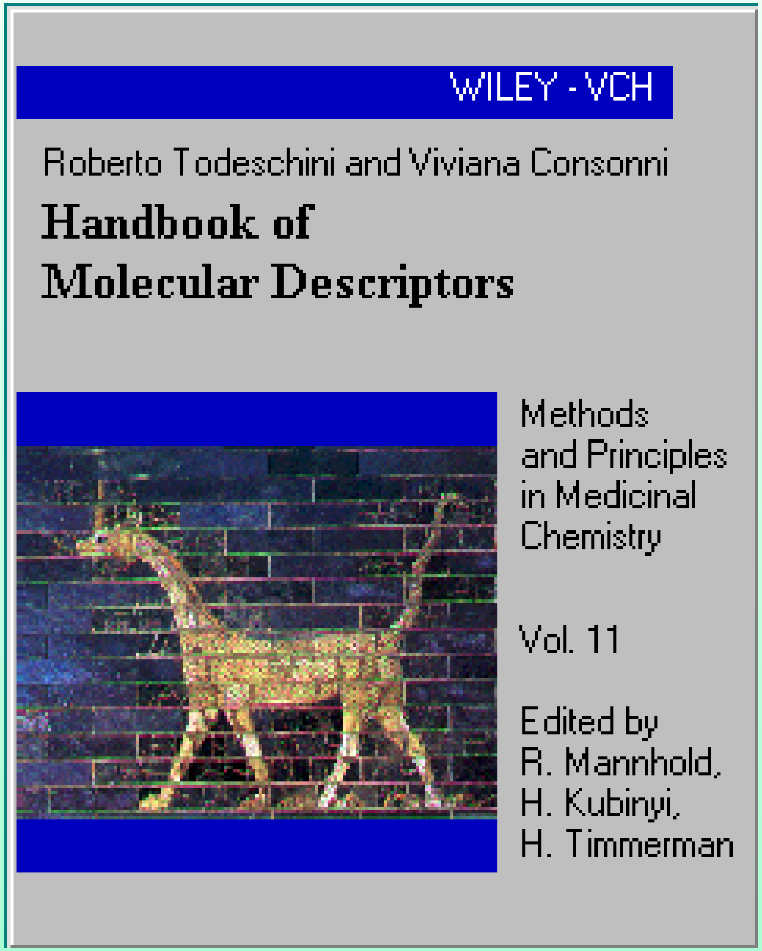
Wilhelm KÖRNER

Molecular descriptors

Definition of molecular descriptor

> "The **molecular descriptor** is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment."
>
> R. Todeschini and V. Consonni

Molecular descriptors



WILEY - VCH

Roberto Todeschini and Viviana Consonni

Handbook of Molecular Descriptors

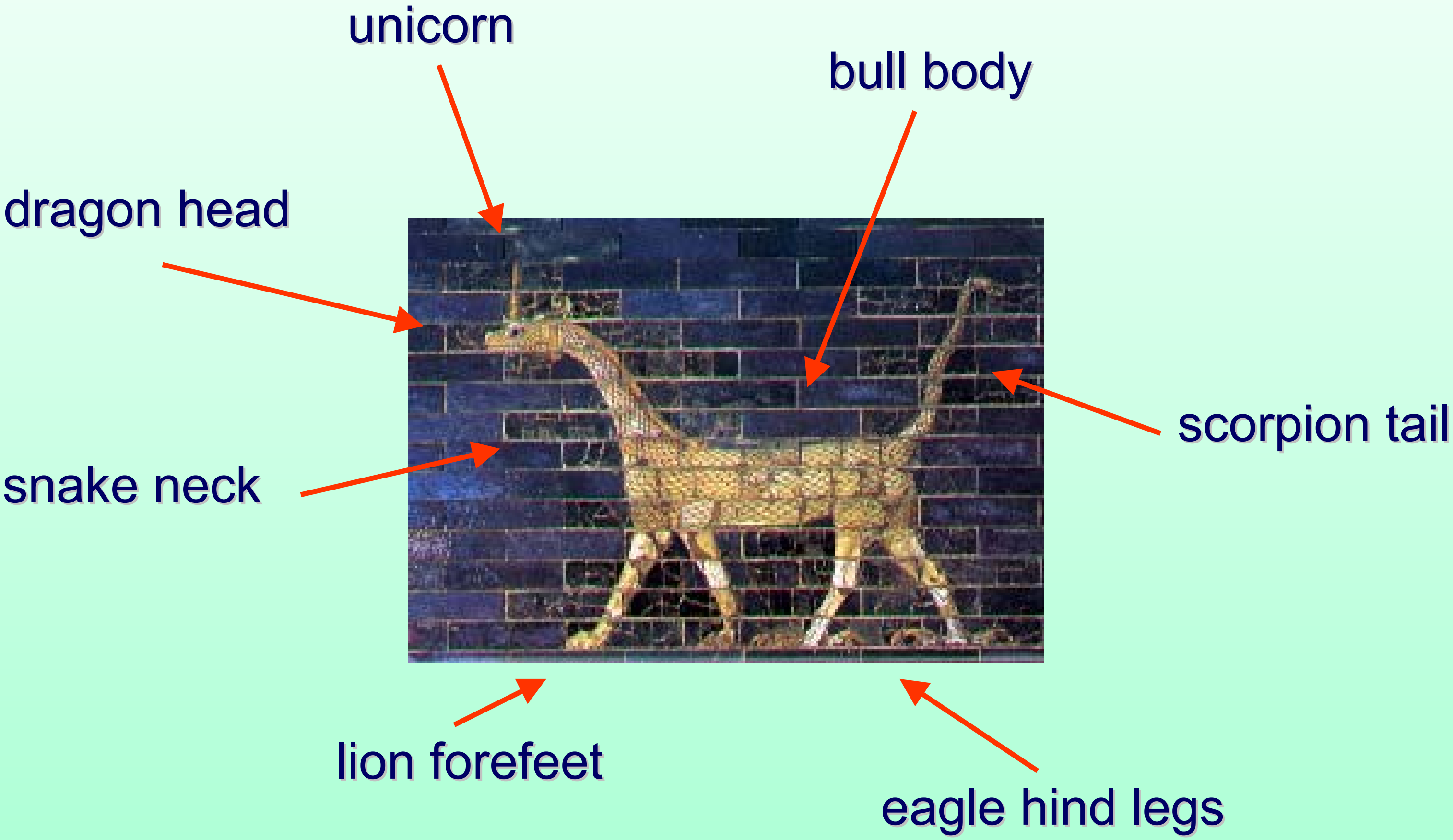
Methods and Principles in Medicinal Chemistry

Vol. 11

Edited by
R. Mannhold,
H. Kubinyi,
H. Timmerman

≈ 3300 molecular descriptors

Molecular descriptors



unicorn

bull body

dragon head

scorpion tail

snake neck

lion forefeet

eagle hind legs

Molecular descriptors

symmetry

electronic aspects

branching

H - bonding

steric

hydrophobicity

size

shape

reactivity

cyclicity

Molecular descriptors

symmetry

electronic aspects

branching

H - bonding

several meanings in just one number

steric

hydrophobicity

size

shape

reactivity

cyclicity

“Molecular Descriptors for Chemoinformatics”

Roberto Todeschini and Viviana Consonni

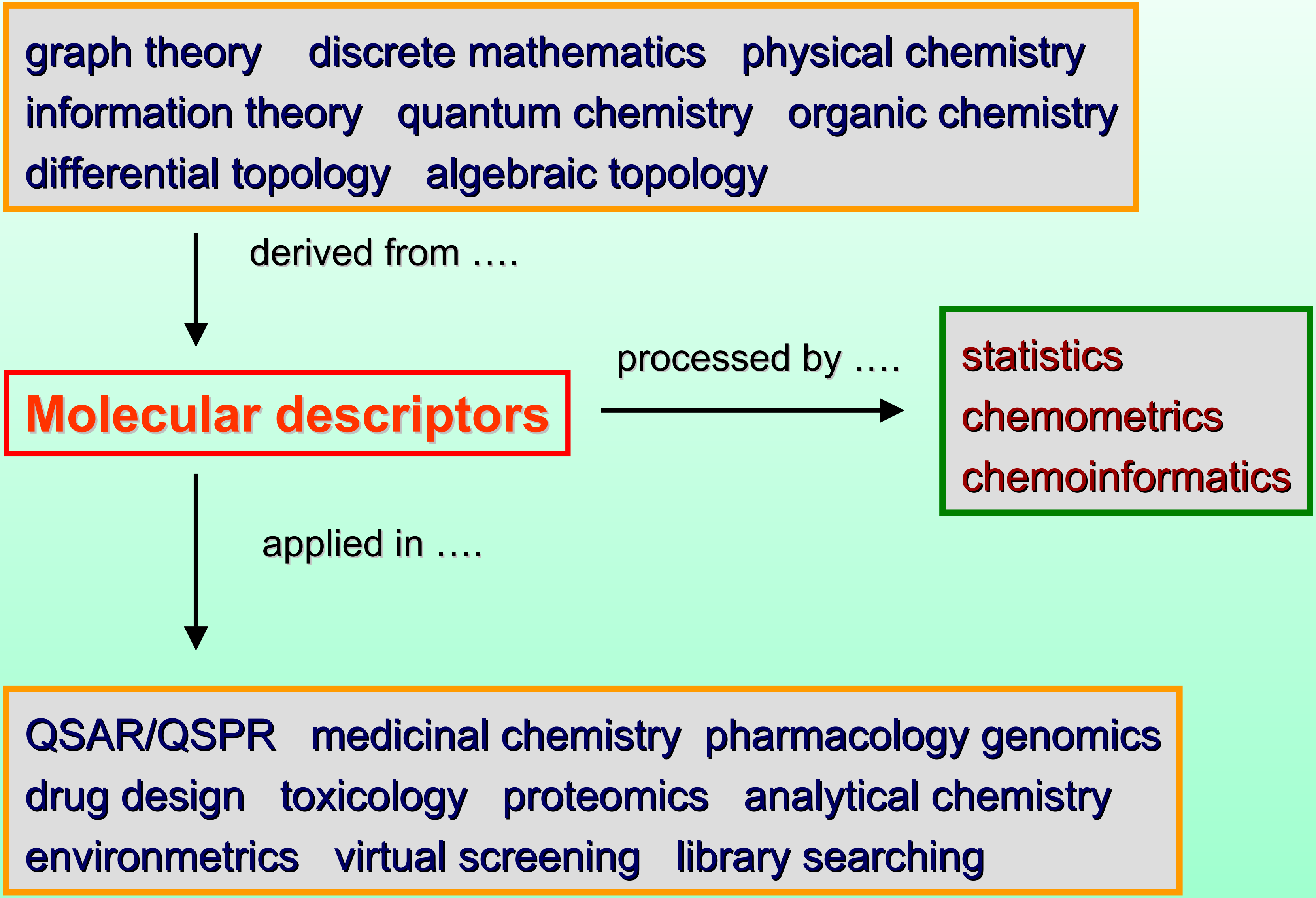
Wiley-VCH

2 volumes

- 6400 bibliographic references
- 1300 pages
- 3000 entries
- 7000 cited authors
- unknown number of formulas

In press

Molecular descriptors



graph theory discrete mathematics physical chemistry information theory quantum chemistry organic chemistry differential topology algebraic topology

derived from ….

Molecular descriptors

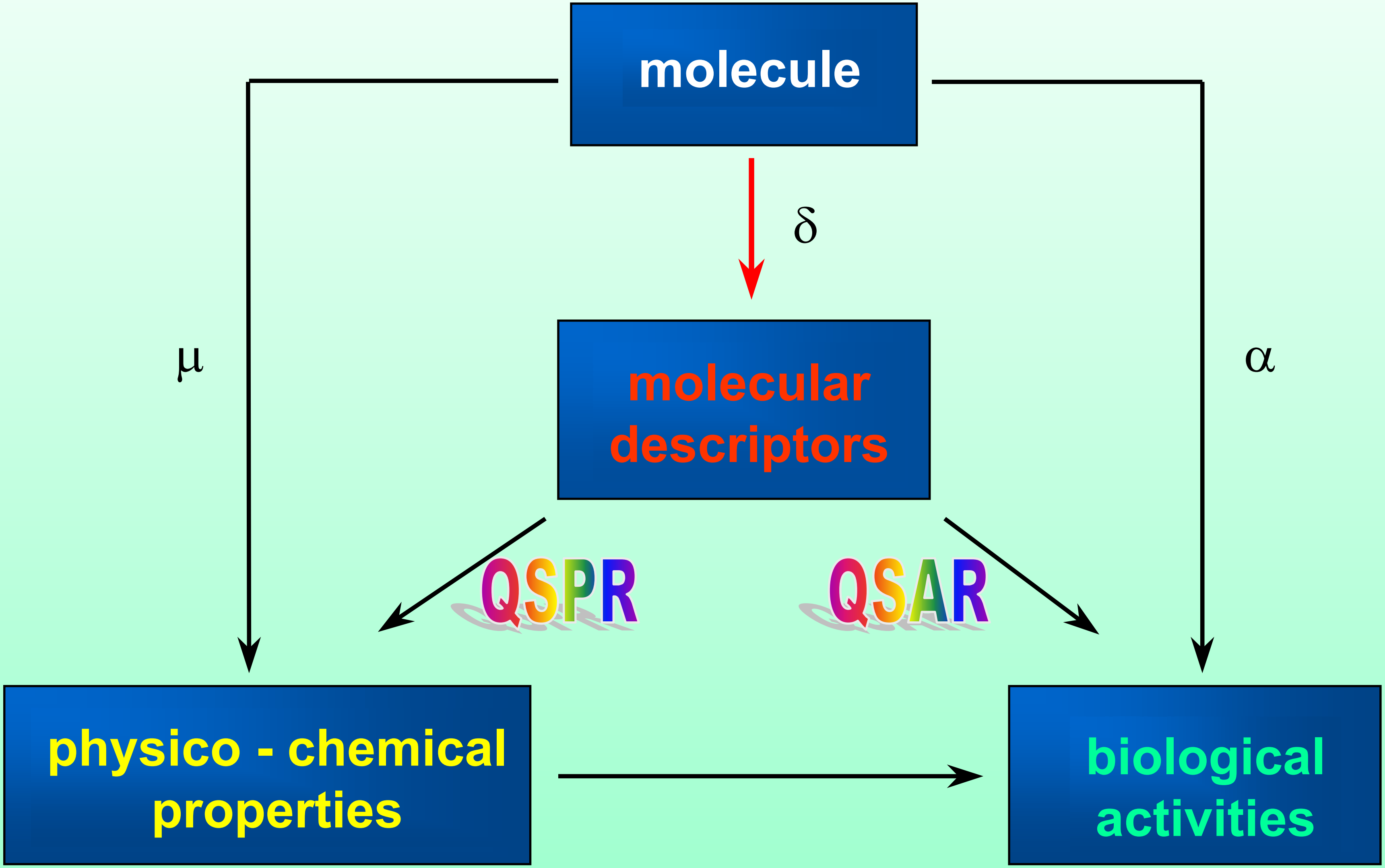
processed by ….

statistics
chemometrics
chemoinformatics

applied in ….

QSAR/QSPR medicinal chemistry pharmacology genomics drug design toxicology proteomics analytical chemistry environmetrics virtual screening library searching

Molecular descriptors



molecule

δ

molecular descriptors

μ

α

QSPR

QSAR

physico - chemical properties

biological activities

The role of the molecular descriptors

Physico-chemical properties

boiling point

melting point

dipole moment

molar refractivity

parachor

octanol/water partition coefficient

vapor pressure

density

solubility

....

The role of the molecular descriptors

Biological activities

binding affinity

lethal dose

inhibition concentration

mutagenicity

carcinogenicity

antiinflammatory activity

antidepressant activity

skin sensitization

....

The role of the molecular descriptors

Environmental properties

biodegradation

bioconcentration

BOD

COD

half - life time

mobility

atmospheric persistance

....

The role of the molecular descriptors

.... and more

conductivity

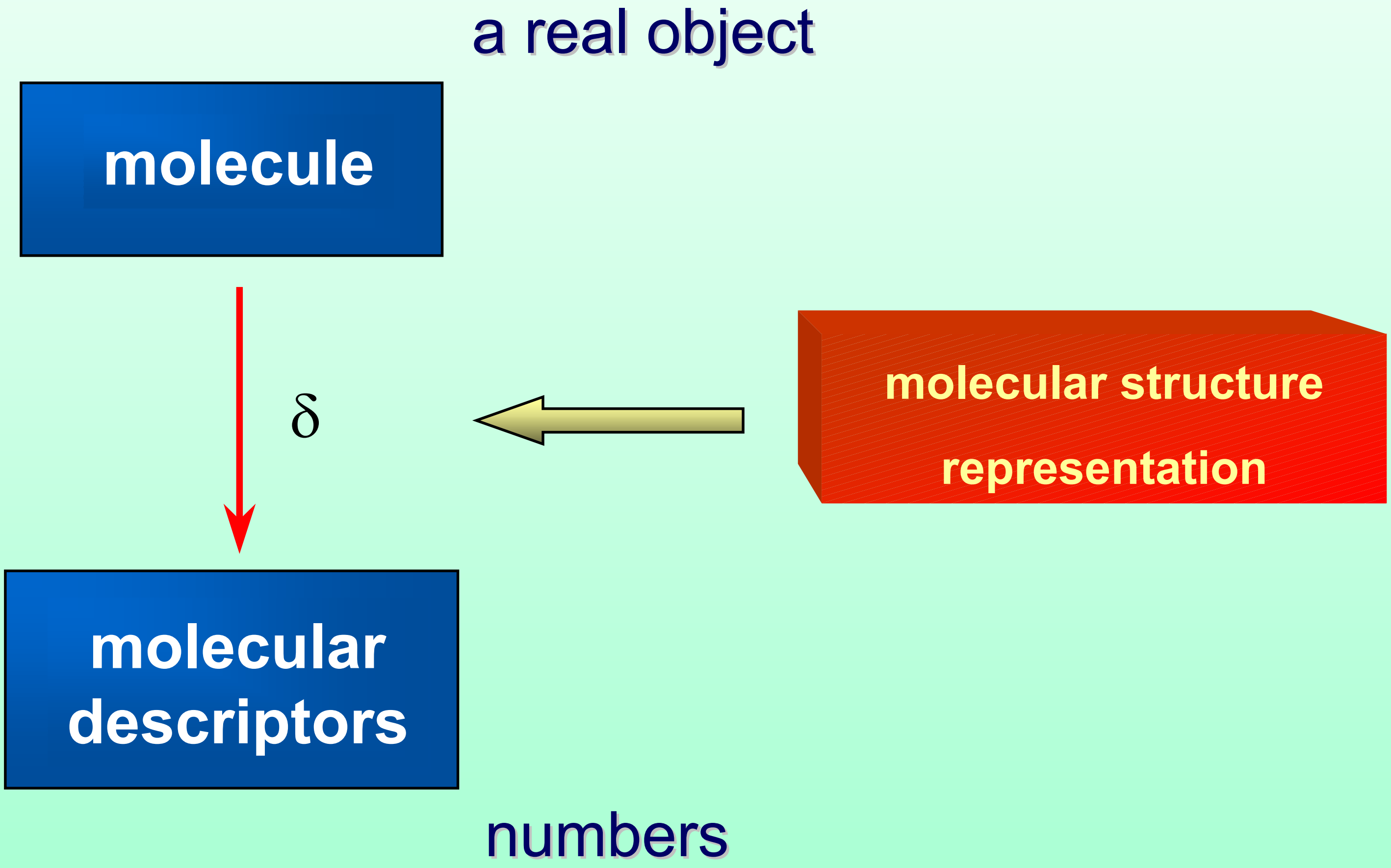
retention time

glass transition temperature

reological behaviours

....

Representations of a molecular structure



a real object

molecule

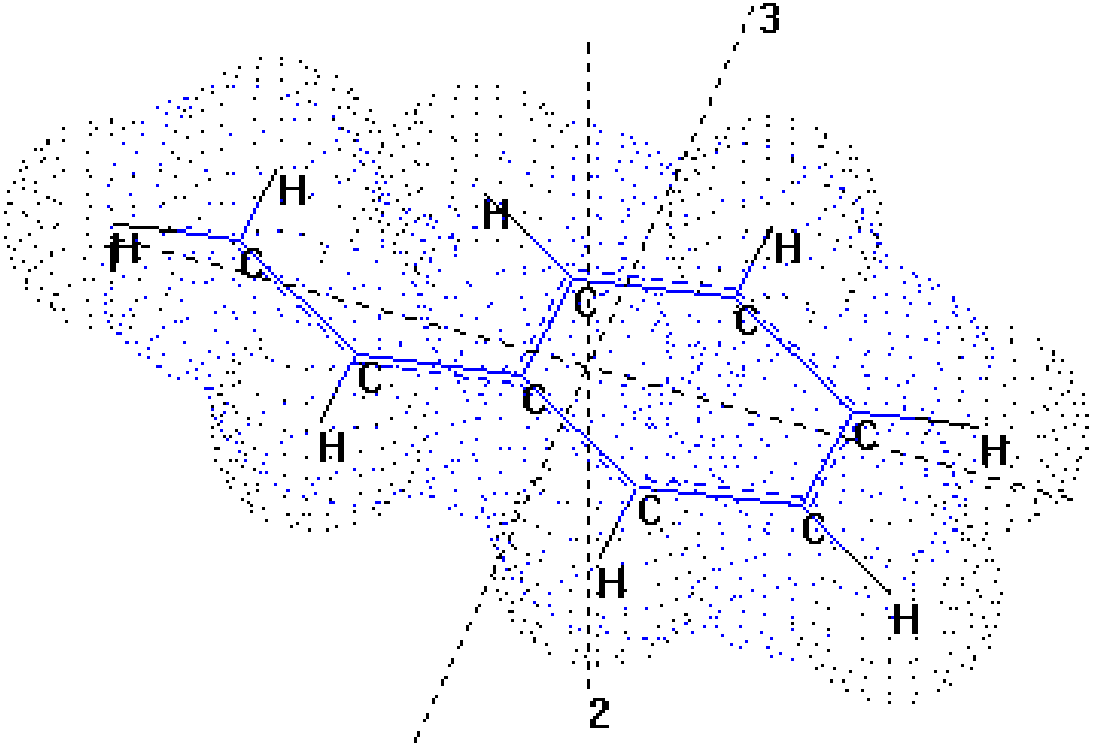
δ

molecular structure representation

molecular descriptors

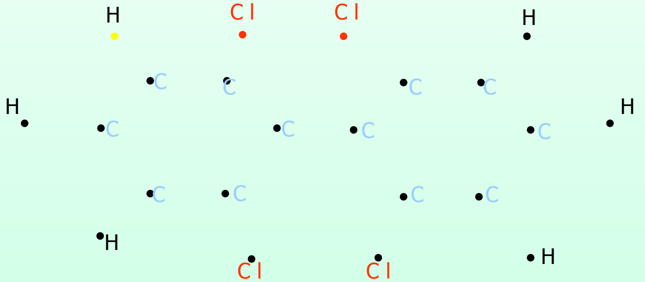
numbers

Representations of a molecular structure

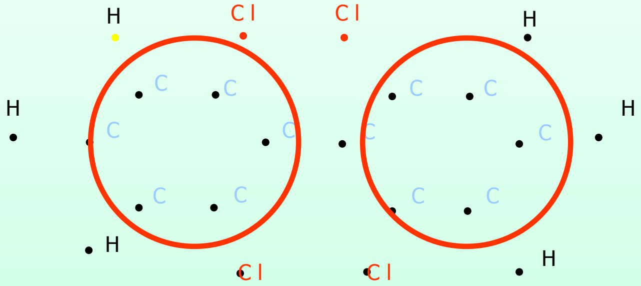


Representations of a molecular structure

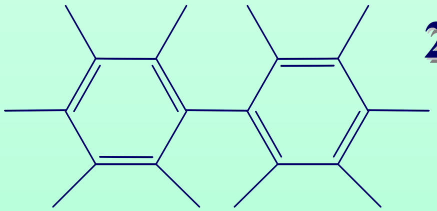
0D - counts



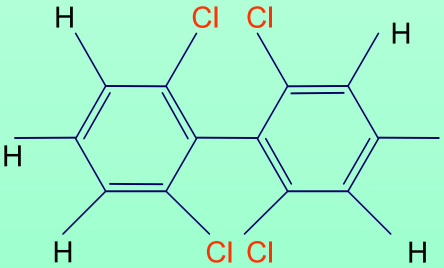
1D – fragment counts



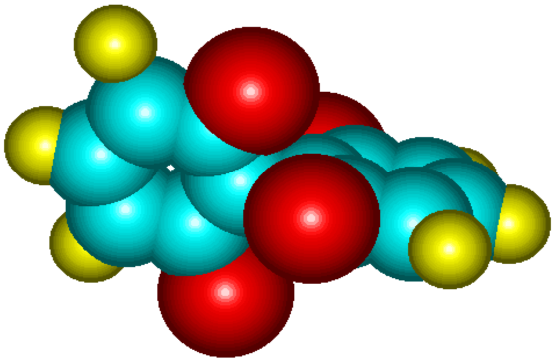
2D - topostructural



2D - topochemical



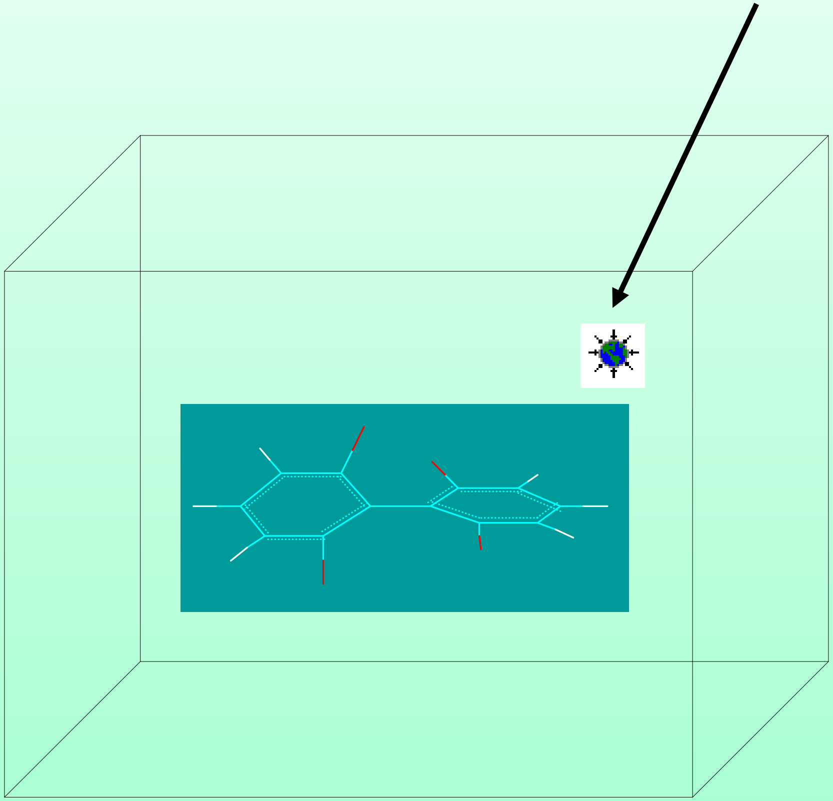
3D - geometrical



Representations of a molecular structure

probes

- steric
- electronic
- hydrophobic



interaction energy value
at each point
for each probe

4D

Properties of a molecular descriptor

Several scientists are involved in searching for new molecular descriptors able to catch new aspects of the molecular structure. This kind of reasearch involves creativity and imagination together with solid theoretical basis allowing to obtain numbers with some structural chemical meaning.

"There are no restriction on the design of structural invariants, the limiting factor is one's own imagination." [1].

M. Randic (1996), *Molecular bonding profiles*, J. Math. Chem., 19, 375-392

Properties of a molecular descriptor

a descriptor MUST have ...

- invariance with respect to labeling and numbering of atoms
- invariance with respect to roto-translation
- an unambiguous algorithmically computable definition
- values in a suitable numerical range for the set of molecules where it is applicable to

Properties of a molecular descriptor

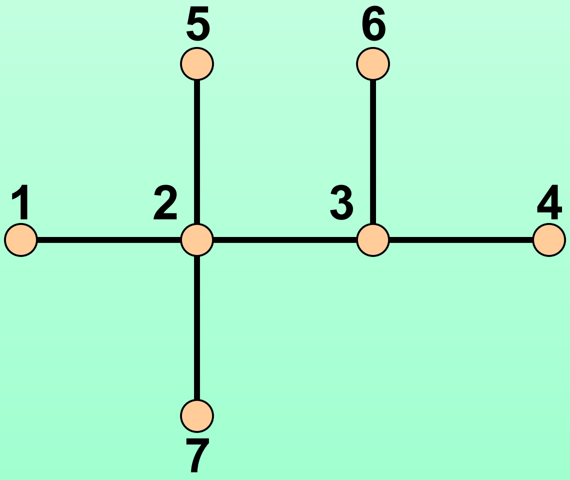
a descriptor should have ...

- a structural interpretation
- a good correlation with at least one property
- no trivial correlation with other molecular descriptors
- gradual change in its values with gradual changes in the molecular structure
- not including in the definition experimental properties
- not restricted to a too small class of molecular structures
- preferably, some discrimination power among isomers
- preferably, not trivially including in the definition other molecular descriptors
- preferably, allowing reversible decoding (back from the descriptor value to the structure)

Molecular descriptors

... some more details about molecular descriptors

Molecular graph



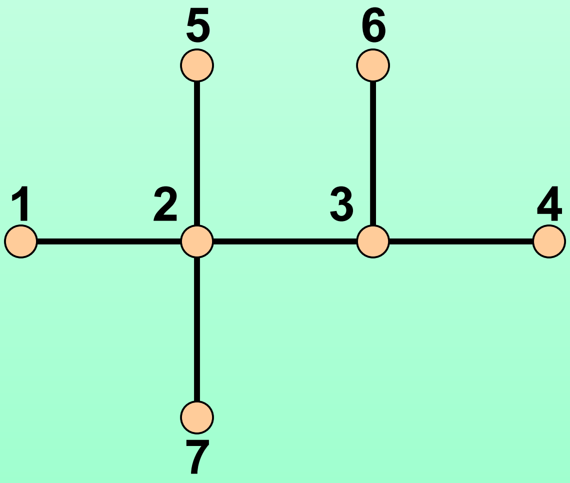
Molecular graph

Mathematical object defined as

$$G = (\mathcal{V}, \mathcal{E})$$

set $\mathcal{V}$ vertices → atoms

set $\mathcal{E}$ edges → bonds



Topological matrices

Adjacency matrix

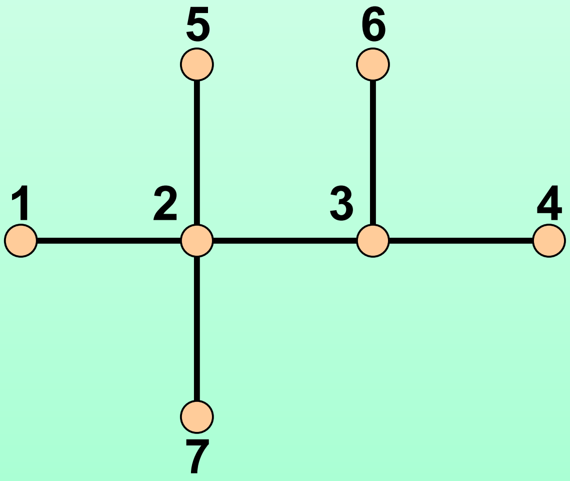
Derived from a molecular graph, it represents the whole set of connections between adjacent pairs of atoms.

$$a_{ij} = \begin{cases} 1 \text{ if atom } i \text{ and } j \text{ are bonded} \\ 0 \text{ otherwise} \end{cases}$$

Local vertex invariants

atom vertex degree

δ_i It is the row sum of the vertex adjacency matrix

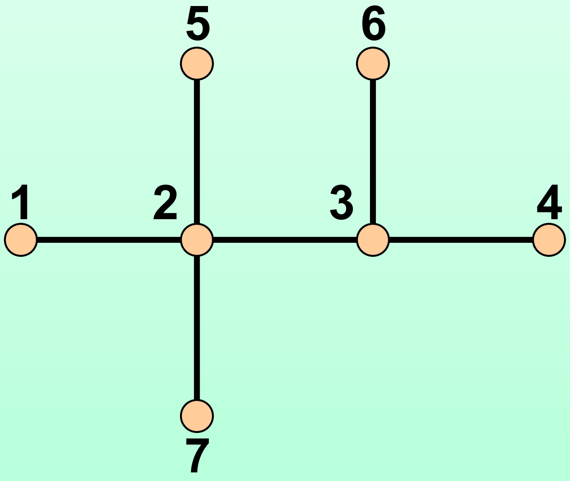


$$\begin{array}{c|ccccccc|c}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & \delta_i \\
\hline
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
2 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 4 \\
3 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 3 \\
4 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
5 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
6 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
7 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}$$

Distance matrix

vertex distance matrix degree

s_i It is the row sum of the vertex distance matrix



	1	2	3	4	5	6	7	s_i	η_i
1	0	1	2	3	2	3	2	13	3
2	1	0	1	2	1	2	1	8	2
3	2	1	0	1	2	1	2	9	2
4	3	2	1	0	3	2	3	14	3
5	2	1	2	3	0	3	2	13	3
6	3	2	1	2	3	0	3	14	3
7	2	1	2	3	2	3	0	13	3

The distance d_{ij} between two vertices is the smallest number of edges between them.

s_i is high for terminal vertices and low for central vertices

Strategies for molecular descriptors

From local vertex invariants you can:

1. $\mathcal{D}_1\left(k;\alpha\right)=k\cdot\sum_{i=1}^{A}\mathcal{L}_i^{\alpha}$

2. $\mathcal{D}_2\left(k;\alpha\right)=k\cdot\sum_{i=1}^{A}\sum_{j=1}^{A}\left(\mathcal{L}_i\cdot\mathcal{L}_j\right)^{\alpha}\quad j\neq i$

3. $\mathcal{D}_3\left(k;\alpha\right)=k\cdot\sum_{i=1}^{A}\sum_{j=1}^{A}a_{ij}\cdot\left(\mathcal{L}_i\cdot\mathcal{L}_j\right)^{\alpha}$

4. $\mathcal{D}_4\left(k;\alpha\right)=k\cdot\left(\prod_{i=1}^{A}\mathcal{L}_i\right)^{\alpha}$

5. $\mathcal{D}_5\left(k\right)=k\cdot\max_{i\in A}\left(\mathcal{L}_i\right)$

6. $\mathcal{D}_6\left(k;\alpha;m\right)=k\cdot\sum_{i=1}^{A}\sum_{j=1}^{A}\left(\mathcal{L}_i\cdot\mathcal{L}_j\right)^{\alpha}\cdot\delta\left(d_{ij};m\right)$

7. $\mathcal{D}_7\left(k;\alpha;m\right)=k\cdot\max_{i,j\in A}\left[\left(\mathcal{L}_i\cdot\mathcal{L}_j\right)^{\alpha}\cdot\delta\left(d_{ij};m\right)\right]$

Strategies for molecular descriptors

Molecular matrices from molecular topology:
- adjacency, distance, detour, Laplace, ...

Functions of the basic molecular matrices:
reciprocal, combined, extended,
complementary, weighted, layered, ...

... more than 100!

Strategies for molecular descriptors

From molecular matrices you can:

1. $\mathcal{D}_1 = \frac{1}{2} \cdot \sum_{i=1}^{A} \sum_{j=1}^{A} m_{ij}$

2. $\mathcal{D}_2 = \frac{1}{2} \cdot \sum_{i=1}^{A} \sum_{j=1}^{A} a_{ij} \cdot m_{ij}$

3. $\mathcal{D}_3\left(k\right) = k \cdot \det\left(\mathbf{M}\right)$

4. $\mathcal{D}_4\left(Sp\right) = f\left(Spectrum\right)$

Strategies for molecular descriptors

From the spectrum eigenvalues of a matrix:

$$SpSum^{k}\left(\mathbf{M},w\right)=\sum_{i=1}^{n}\left|\lambda_{i}\right|^{k} \qquad SpSum_{+}^{k}\left(\mathbf{M},w\right)=\sum_{i=1}^{n^{+}}\left(\lambda_{i}^{+}\right)^{k} \qquad SpSum_{-}^{k}\left(\mathbf{M},w\right)=\sum_{i=1}^{n^{-}}\left|\lambda_{i}^{-}\right|^{k}$$

$$SpAD\left(\mathbf{M},w\right)=\sum_{i=1}^{n}\left|\lambda_{i}-\bar{\lambda}\right| \qquad SpMAD\left(\mathbf{M},w\right)=\sum_{i=1}^{n}\left|\lambda_{i}-\bar{\lambda}\right|/n$$

$$MinSp\left(\mathbf{M},w\right)=\min_{i}\left\{\lambda_{i}\right\} \qquad MaxSp\left(\mathbf{M},w\right)=\max_{i}\left\{\lambda_{i}\right\}$$

$$MaxSpA\left(\mathbf{M},w\right)=\max_{i}\left\{\left|\lambda_{i}\right|\right\} \qquad SpDiam\left(\mathbf{M},w\right)=MaxSp\text{ - }MinSp$$

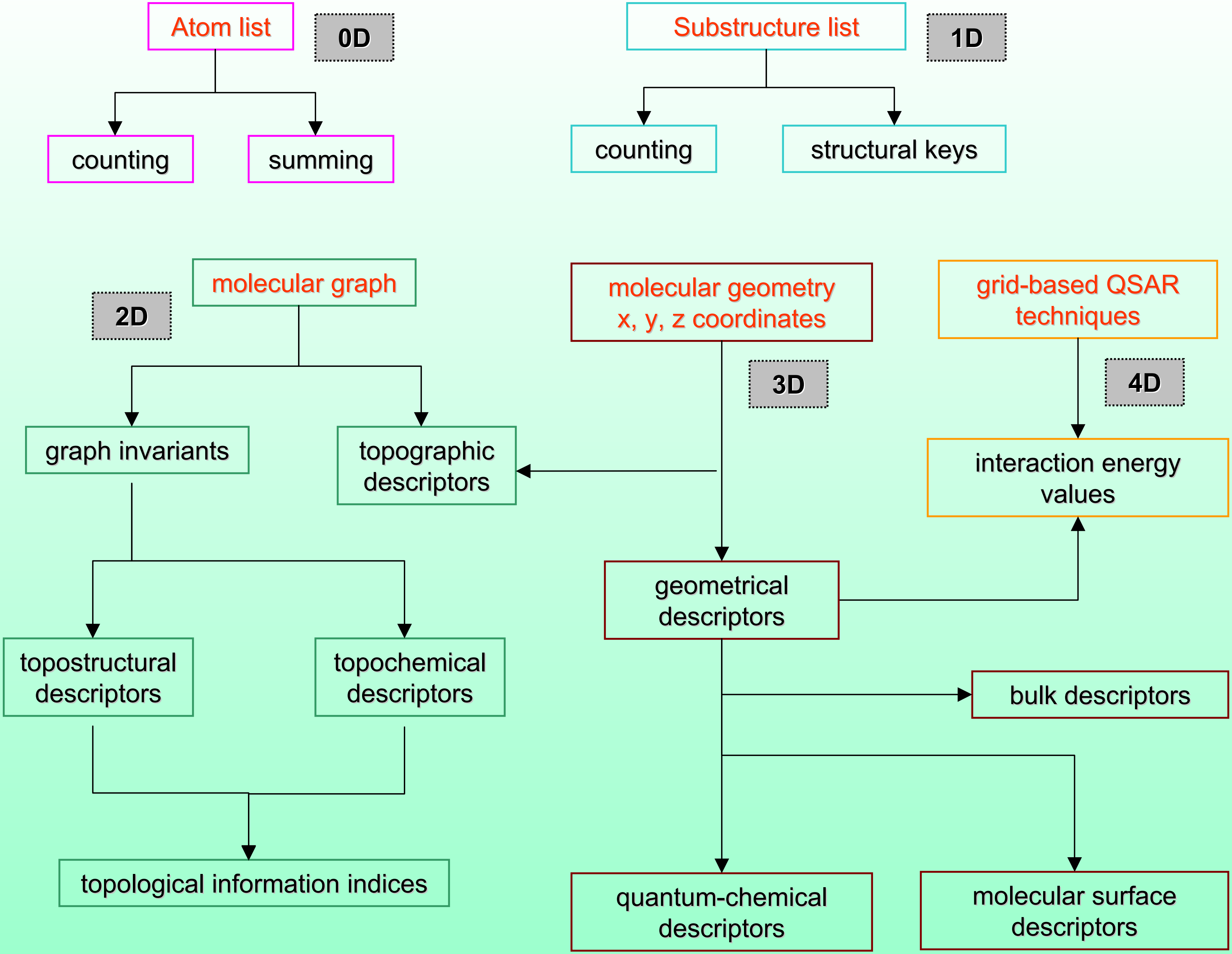
Strategies for molecular descriptors

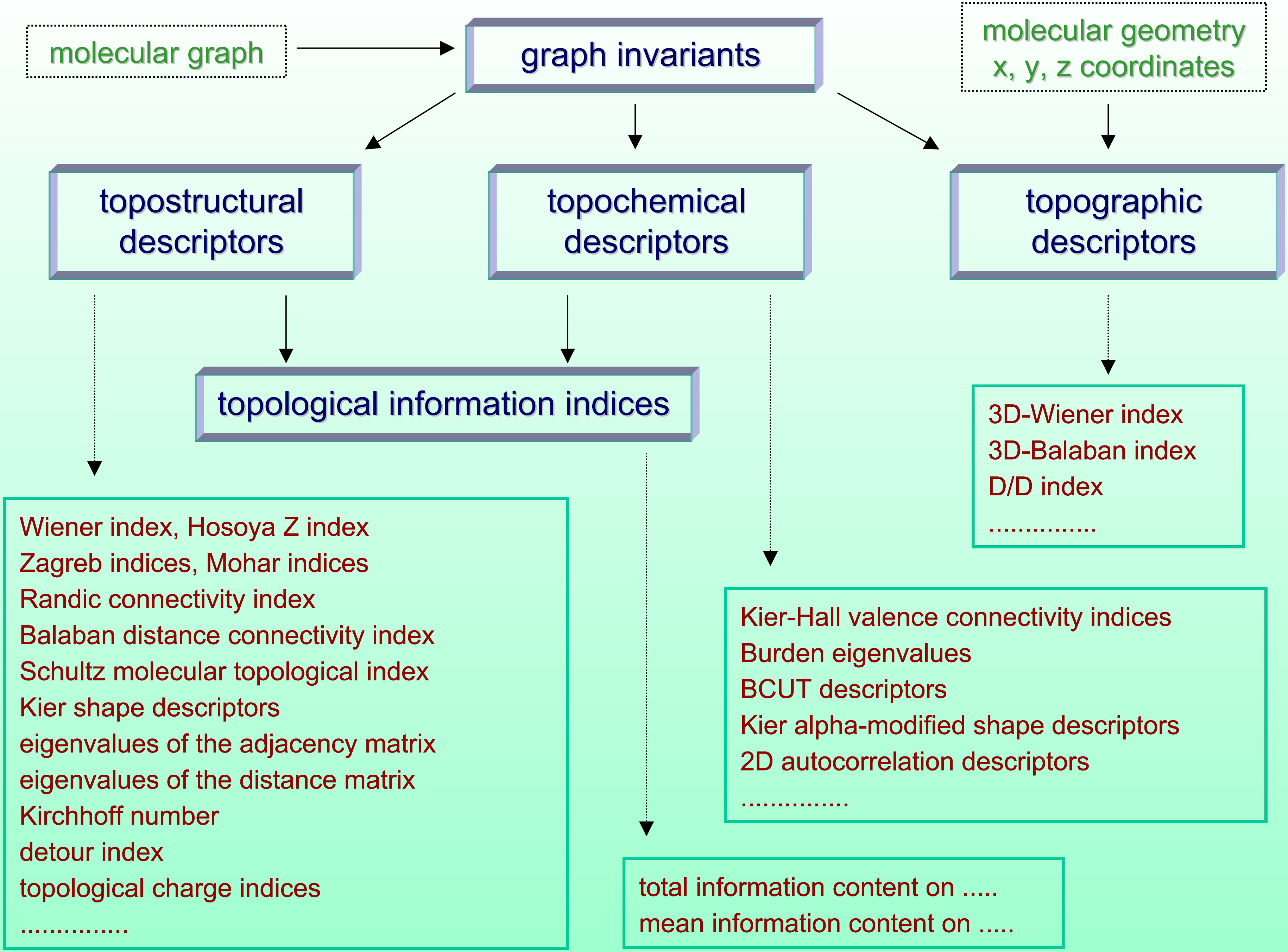
3D atom coordinates and geometry matrix:

$$\mathbf{M} = \begin{vmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \dots & \dots & \dots \\ x_A & y_A & z_A \end{vmatrix} \Rightarrow \mathbf{G} \equiv \begin{vmatrix} 0 & r_{12} & \dots & r_{1A} \\ r_{21} & 0 & \dots & r_{2A} \\ \dots & \dots & \dots & \dots \\ r_{A1} & r_{A2} & \dots & 0 \end{vmatrix}$$

$$\Downarrow$$

... a lot of new local invariants and 3D molecular descriptors are derived !


Atom list
0D
counting
summing
Substructure list
1D
counting
structural keys
molecular graph
2D
graph invariants
topographic descriptors
topostructural descriptors
topochemical descriptors
topological information indices
molecular geometry x, y, z coordinates
3D
geometrical descriptors
bulk descriptors
quantum-chemical descriptors
molecular surface descriptors
grid-based QSAR techniques
4D
interaction energy values



molecular graph

graph invariants

molecular geometry
x, y, z coordinates

topostructural descriptors

topochemical descriptors

topographic descriptors

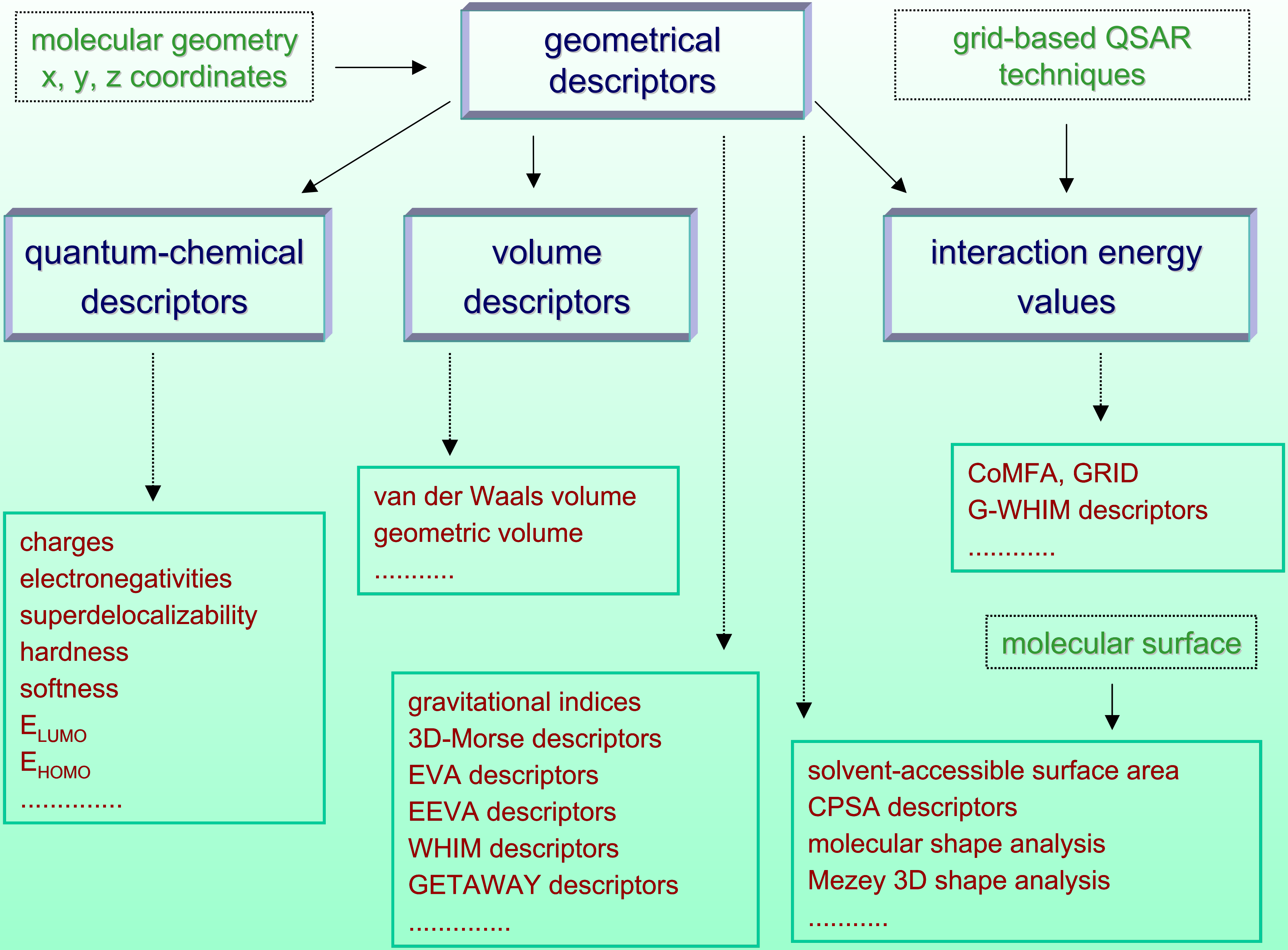
topological information indices

3D-Wiener index
3D-Balaban index
D/D index
....

Wiener index, Hosoya Z index
Zagreb indices, Mohar indices
Randic connectivity index
Balaban distance connectivity index
Schultz molecular topological index
Kier shape descriptors
eigenvalues of the adjacency matrix
eigenvalues of the distance matrix
Kirchhoff number
detour index
topological charge indices
....

Kier-Hall valence connectivity indices
Burden eigenvalues
BCUT descriptors
Kier alpha-modified shape descriptors
2D autocorrelation descriptors
....

total information content on .....
mean information content on .....



molecular geometry
x, y, z coordinates

geometrical descriptors

grid-based QSAR techniques

quantum-chemical descriptors

- charges
- electronegativities
- superdelocalizability
- hardness
- softness
- E_{LUMO}
- E_{HOMO}
- ....

volume descriptors

- van der Waals volume
- geometric volume
- ....

- gravitational indices
- 3D-Morse descriptors
- EVA descriptors
- EEVA descriptors
- WHIM descriptors
- GETAWAY descriptors
- ....

interaction energy values

- CoMFA, GRID
- G-WHIM descriptors
- ....

molecular surface

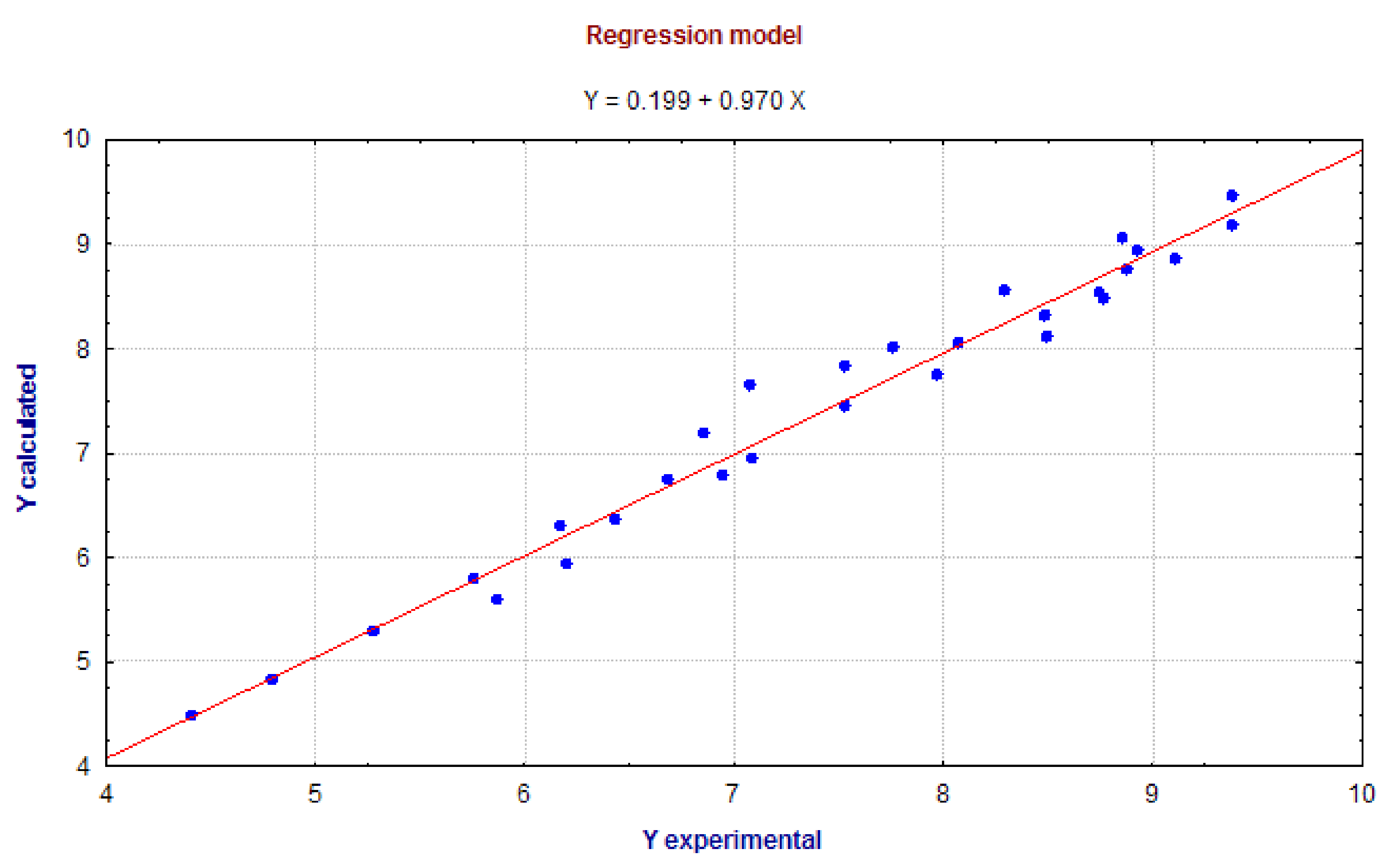
- solvent-accessible surface area
- CPSA descriptors
- molecular shape analysis
- Mezey 3D shape analysis
- ....

QSAR strategy

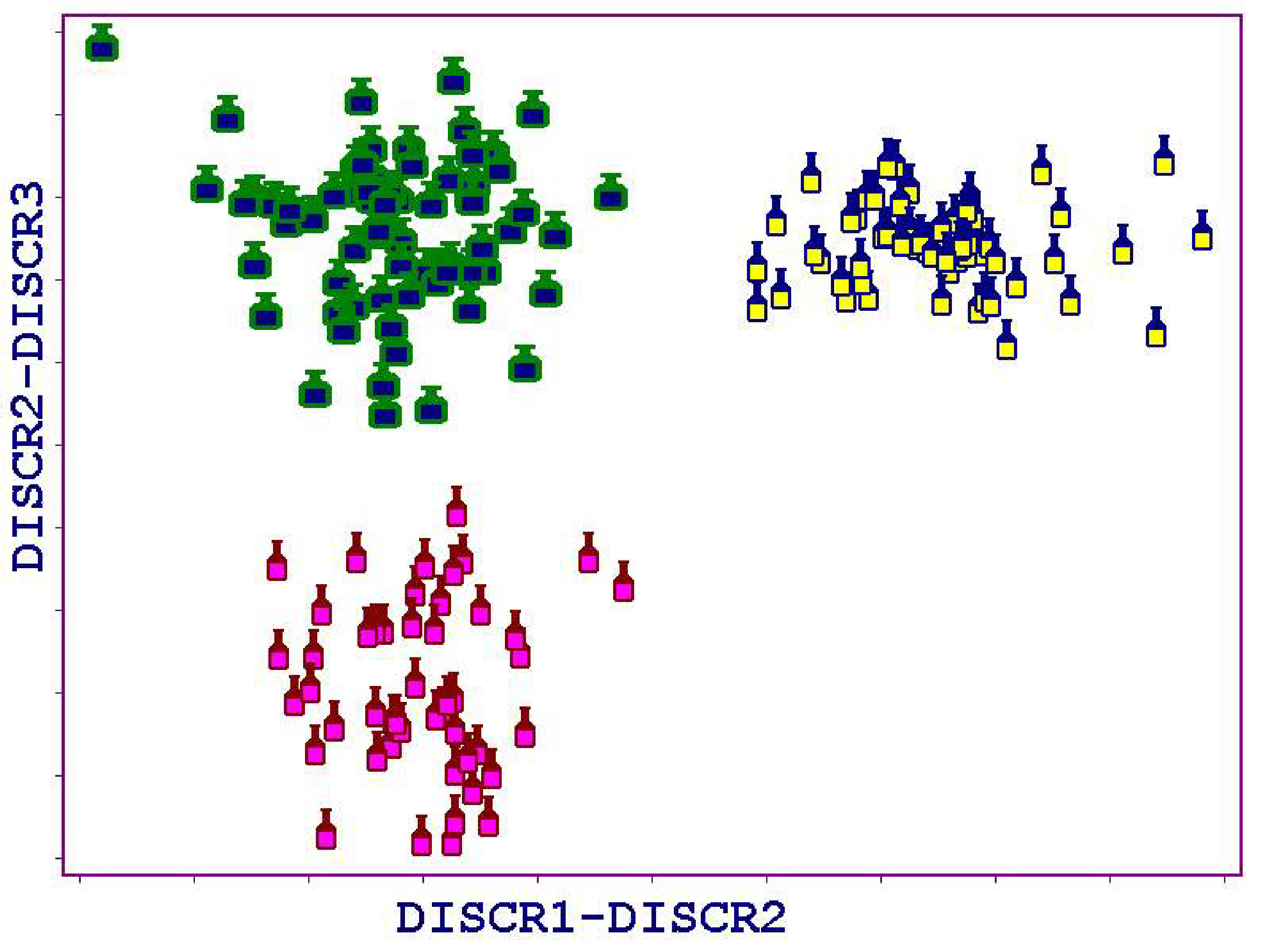
models ...

- regression models (quantitative response)
- classification models (qualitative response)
- ranking models (ordered response)

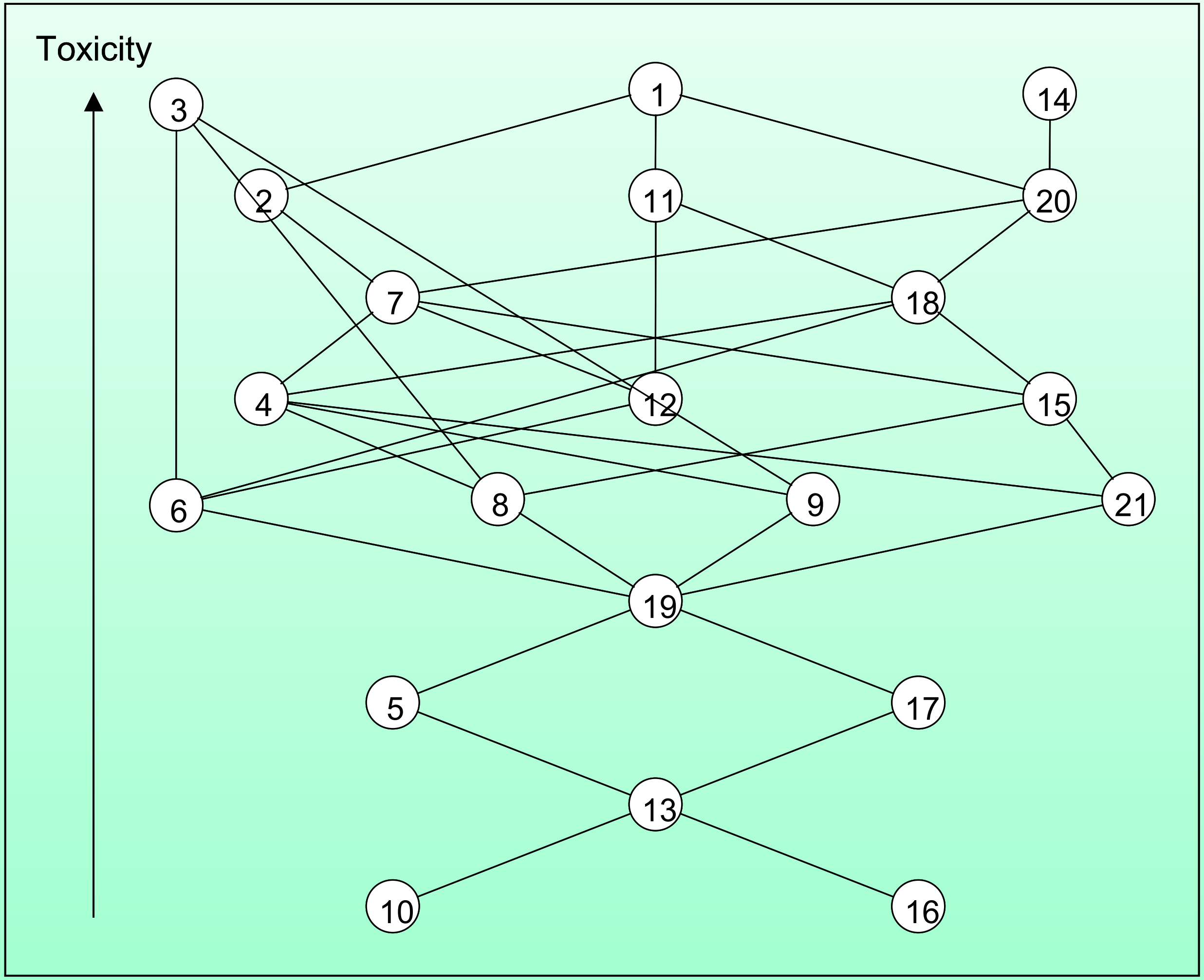
QSAR strategy - Regression



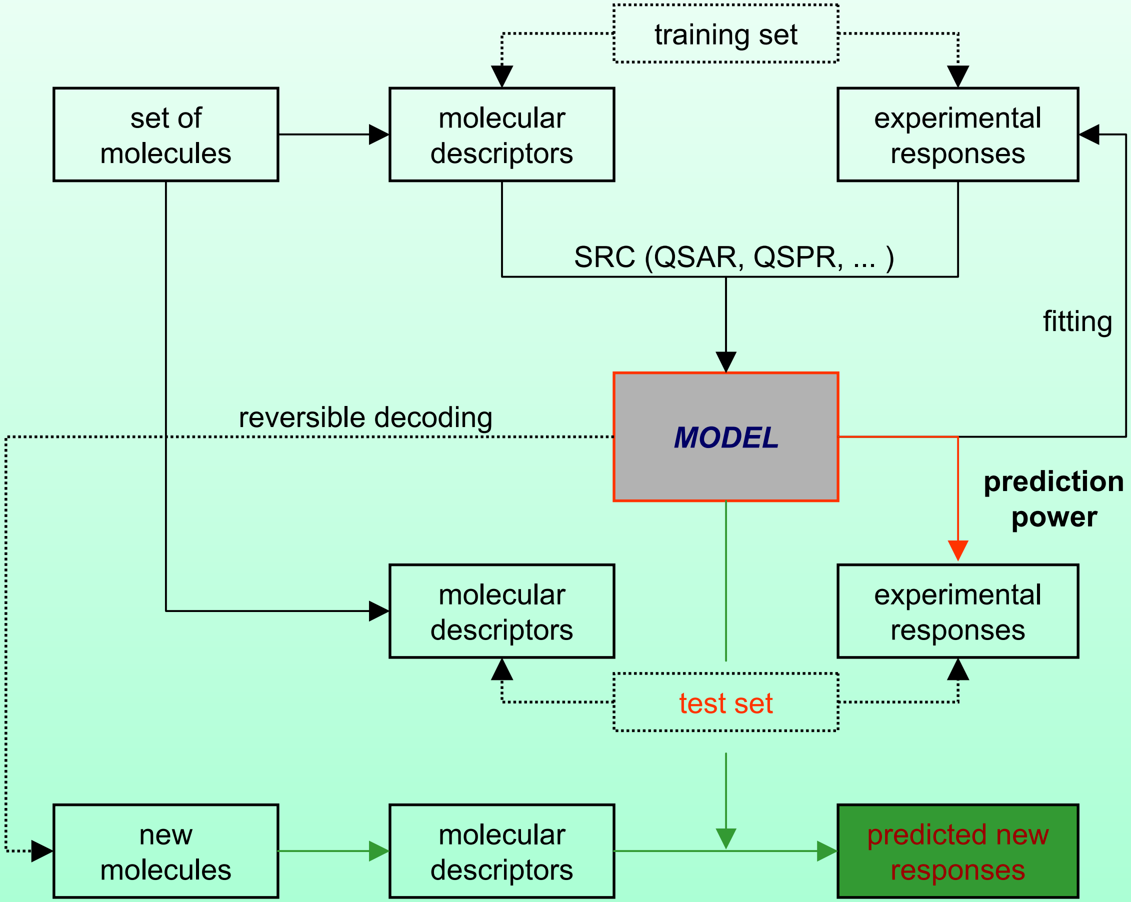
QSAR strategy - Classification



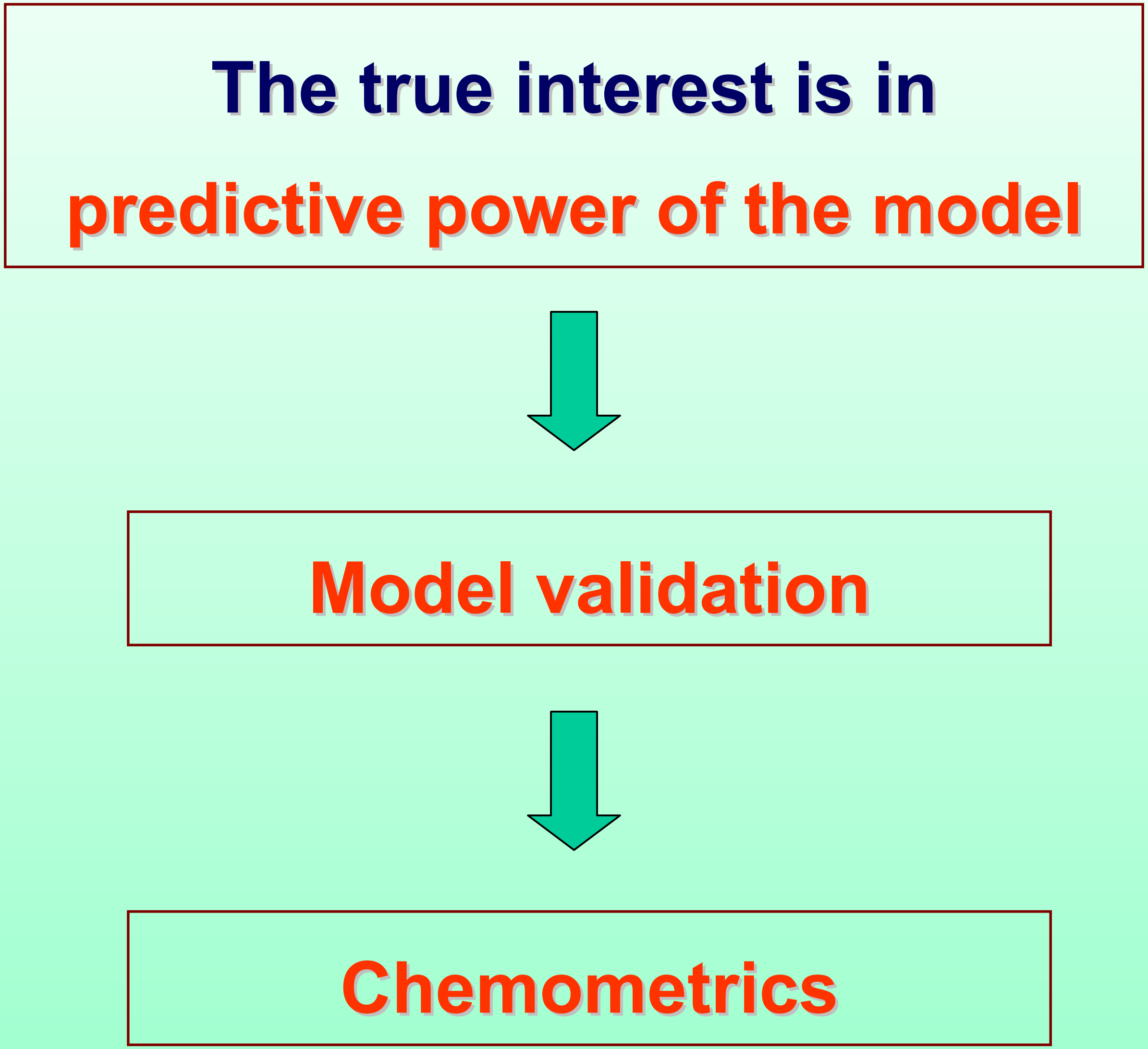
QSAR strategy - Ranking



QSAR strategy



QSAR strategy



The true interest is in

predictive power of the model

Model validation

Chemometrics

FAQ - Frequently Asked Questions

1. What is the meaning of that descriptor ?

2. Why are there some models with the same prediction power but different molecular descriptors ?

3. Why use a huge number of molecular descriptors ?

4. Is a model explaining the known facts of a system better than a model predicting the future events of that system ?

FGA - our Frequently Given Answers

1. What is the meaning of that descriptor ?

A molecular descriptor is a number extracted by a well defined algorithm from a molecular representation of a complex system, i.e. the molecule. There are good reasons to believe that often our difficulties to attribute a meaning to this number ultimately flow from the lacking of deeper chemical theories and higher level languages and not from exoteric approaches to the descriptor definition.

R. Todeschini and V. Consonni

FGA - our Frequently Given Answers

2. Why are there some models with the same prediction power but different molecular descriptors ?

Molecular descriptors are often intercorrelated, therefore different molecular descriptors can, in turn, take part in a model.

↓

Any alternative viewpoint with a different emphasis leads to an inequivalent description. There is only one reality but there are many points of view.

Hans Primas

FGA - our Frequently Given Answers

3. Why use a huge number of molecular descriptors ?

Complexity is not an intrinsic property of systems, but rather arises from the number of ways in which we are able (or desire) to interact with a system.

↓

A molecule is undoubtedly a complex system

FGA - our Frequently Given Answers

4. Is a model explaining the known facts of a system better than a model predicting the future events of that system ?

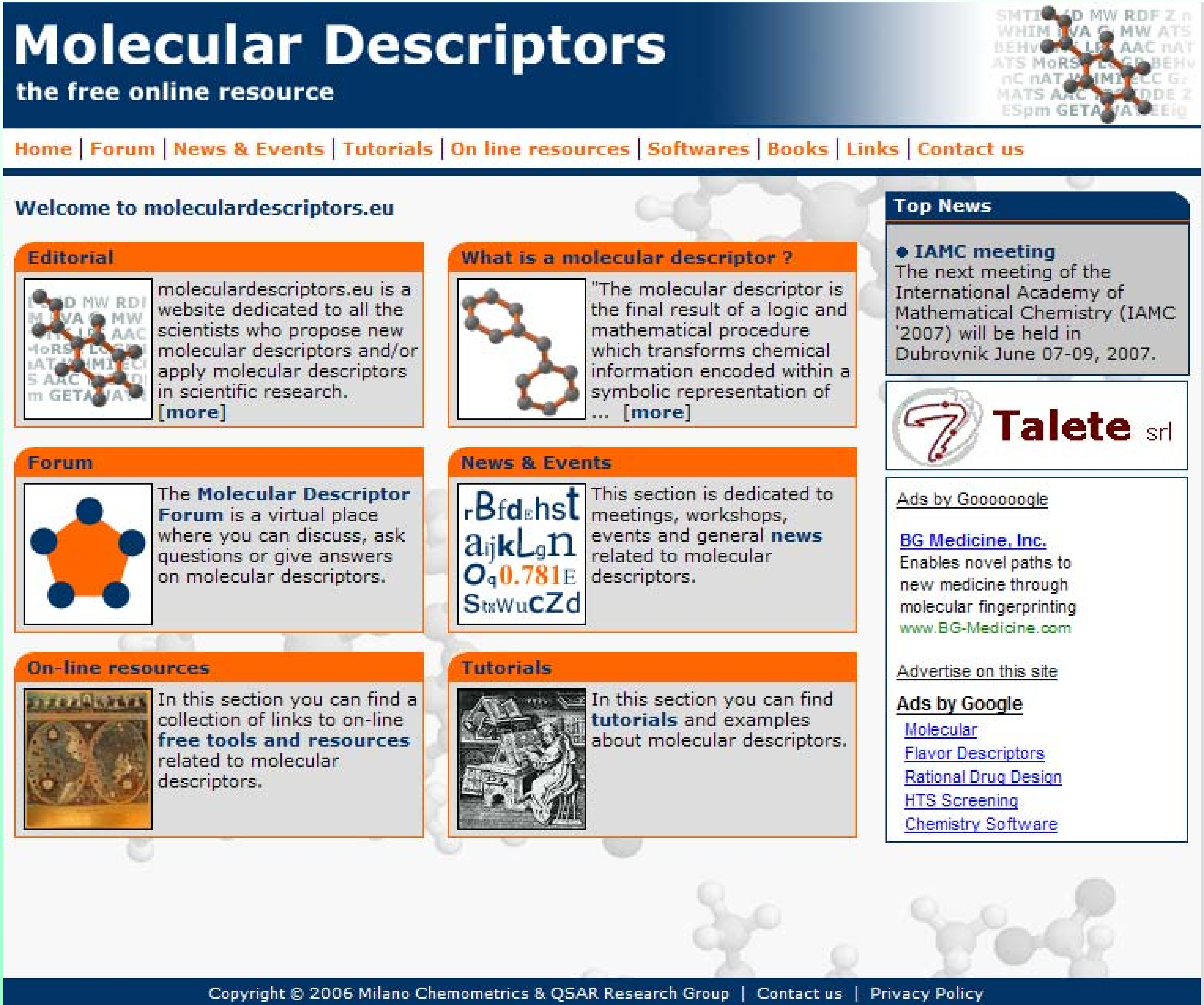
Don’t forget your goal!

An understanding of the behavior of a system does not always coincide with the prediction of the system’s future behavior!

↓

fitting versus prediction

www.moleculardescriptors.eu



Molecular Descriptors

the free online resource

Home | Forum | News & Events | Tutorials | On line resources | Softwares | Books | Links | Contact us

Welcome to moleculardescriptors.eu

Editorial

moleculardescriptors.eu is a website dedicated to all the scientists who propose new molecular descriptors and/or apply molecular descriptors in scientific research. [more]

What is a molecular descriptor ?

"The molecular descriptor is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of ... [more]

Forum

The **Molecular Descriptor Forum** is a virtual place where you can discuss, ask questions or give answers on molecular descriptors.

News & Events

This section is dedicated to meetings, workshops, events and general **news** related to molecular descriptors.

On-line resources

In this section you can find a collection of links to on-line **free tools and resources** related to molecular descriptors.

Tutorials

In this section you can find **tutorials** and examples about molecular descriptors.

Top News

● **IAMC meeting**
The next meeting of the International Academy of Mathematical Chemistry (IAMC '2007) will be held in Dubrovnik June 07-09, 2007.

Milano Chemometrics and QSAR Research Group

Prof. Roberto Todeschini
Dr. Viviana Consonni
Dr. Manuela Pavan
Dr. Andrea Mauri
Dr. Davide Ballabio
Dr. Alberto Manganaro

chemometrics
molecular descriptors
QSAR
multicriteria decision making
environmetrics
experimental design
artificial neural networks
statistical process control

Department of Environmental Sciences
University of Milano - Bicocca
P.za della Scienza, 1 - 20126 Milano (Italy)
Website: www.disat.unimib.it/chm/

Milano Chemometrics and QSAR Research Group

Prof. Roberto Todeschini
Dr. Viviana Consonni
Dr. Manuela Pavan
Dr. Andrea Mauri
Dr. Davide Ballabio
Dr. Alberto Manganaro

Department of Environmental Sciences
University of Milano - Bicocca
P.za della Scienza, 1 - 20126 Milano (Italy)
Website: www.disat.unimib.it/chm/

THANK YOU