Research Signpost
37/661 (2), Fort P.O.
Trivandrum-695 023
Kerala, India

QSPR-QSAR Studies on Desired Properties for Drug Design, 2010: 117-165 ISBN: 978-81-308-0404-0
Editor: Eduardo A. Castro

5. From molecular structure to molecular design through the Molecular Descriptors Family Methodology

Sorana D. Bolboacă[1] and Lorentz Jäntschi[2]

[1]"Iuliu Hatieganu" University of Medicine and Pharmacy Cluj-Napoca, 6 Louis Pasteur 400349 Cluj-Napoca, Romania; [2]Technical University of Cluj-Napoca, 103-105 Muncii Bvd 400641 Cluj-Napoca, Romania

Abstract. The Molecular Descriptors Family on the Structure-Property/Activity Relationships (MDF SPR/SAR) is an area of computational research able to generate a family of molecular descriptors and to build models in order to estimate and predict the property/activity of chemical compounds. This review aims to briefly present the MDF SPR/SAR methodology and to discuss its abilities to estimate and predict different properties and activities.

Introduction

Structure-Activity Relationships (SARs), Structure-Property Relationships (SPRs) and Property-Activity Relationships (PARs) were first introduced by Louis Pluck HAMMETT in 1937 [1]. A more recent review summarizes the most important applications of Hammett's equation [2].

The idea of linking the structure of a compound with its activity or property was published before the introduction of the SAR concept. In 1868,

Correspondence/Reprint request: Dr. Sorana D. Bolboacă, "Iuliu Hatieganu" University of Medicine and Pharmacy Cluj-Napoca, 6 Louis Pasteur, 400349 Cluj-Napoca, Romania. E-mail: sbolboaca@umfcluj.ro

Crum-Brown and Fraser noted that the activity of a compound is a function of its chemical composition and structure [3]. In 1893, Richet and Seancs demonstrated that cytotoxicity was inversely related with water solubility on a set of organic compounds [4]. In 1899, Mayer demonstrated that the narcotic action of a sample of organic compounds was related with solubility in olive oil [5]. Therefore, Hammett [6] and Taft [7] could be said to have laid the basis of QSAR/QSPR development.

Quantitative relationships (QSAR, QSPR, QPAR), which are mathematical approaches to the link between the structure and the property/activity of chemical compounds in a quantitative manner [8], are applied when the property and/or activity are quantitative. Note that not all the properties and activities of chemical compounds can be classified as quantitative. An example could be the sweetness of sugar (one of the five basic tastes, being almost universally perceived as a pleasurable experience), which can be appreciated only through comparison (relative scale) since there is no single reference scale (such as the boiling and freezing point and the Celsius scale for temperature). Properties that are expressed quantitatively may have several reference scales. Consequently, the use of terms such as QSAR, QSPR, and QPAR is currently avoided, the terms (Q)SAR, (Q)SPR, and (Q)PAR, or simply SAR, SPR, and PAR being preferred.

As far as the structure is concerned, things are relatively simpler. Thus, an atom or a bond from a molecule could exist (highlighted through electronic transitions and/or molecular vibrations and/or rotations) or could not exist (0 or 1). Molecular geometry (especially the liquid or gas phase) is more complicated. The Heisenberg principle (Werner HEISENBERG, 1901-1976, one of the founders of quantum mechanics, a Nobel prize winner) demonstrates the rules of uncertainty through the principles of uncertainty (molecular and atomic level) at micro level. Moreover, molecular geometry depends on the environment on which a molecule is placed (the vicinity of the molecule), the temperature, the pressure, etc. Thus, dealing with molecular geometry is a matter of relativity if not a matter of uncertainty. In conclusion, in the field of Structure-Property-Activity Relationships (SPARs) there are certainties (e.g. molecular topology), uncertainties (e.g. molecular geometry), relativities (e.g. biological activities), and evidences (e.g. physical-chemical properties).

Mathematical Chemistry [9,10], Quantum Chemistry [11], and Medicinal Chemistry [12] have increasingly significant contributions to the design and improvement of drugs. The dynamics of pharmaceuticals is high; new drugs appear on the market daily, even if the process is a long-lasting one. Drug design has recently emerged as a new field [13,14]. The usage of computing power was a major breakthrough. Grassy et al. [15] reported a search for

peptides possessing immunosuppressive activity by using 27 descriptors derived from the structure (topology and shape). Twenty-six peptides were selected from a combinatorial library of about 280000 compounds. The predicted activity was high. Five of them were actually synthesized and tested experimentally. The most potent compounds had shown an immunosuppressive activity approximately 100 times higher than the lead compound. Combinatorial Chemistry is now a new field [16,17].

Almost 40 years have passed [18] since QSAR (Quantitative Structure-Activity Relationship) paradigm proved its utility in agriculture, pharmacy, toxicology, and other fields. Many methods (CoMFA [19] - Comparative Molecular Field Analysis and its variants CoMSIA, MSIA - Molecular Similarity Indices Analysis), WHIM [20] - Weighted Holistic Invariant Molecular (and its variant MS-WHIM - Molecular Surface WHIM) MTD [21] - Minimal Topological Distance (and its variant MSD, S - Steric), FPIF [22] - Fragmental Property Index Family, MDF [23] - Molecular Descriptors Family) were introduced and proved to be good estimators and predictors. Other emergent results such as S(Q)SAR (Spectral (Quantitative) Structure Activity Relationship) [24] are again challenging the classical approaches.

The results in cellular genetics, the mechanic dynamics of cells, genetic mutations, sequencing and coding of macromolecular topology information lead to the recording and storing of molecular geometry into databases for a large number of biologically active chemical compounds [25]. Information stored in databases integrated with proper combinatorial methods may lead to the identification of active compounds with high potency.

Statistical methods for internal and external validation [26], correlated correlation analysis [27], and principal component analysis [28] are methods which have a role in identifying (Q)SA/PRs (see [29,30,31] for meaning of parenthesis Q). These methods allow the selection of compounds with the best biological activity.

The aim of the present review is to present the Molecular Descriptors Family (MDF) methodology and the obtained results by applying this methodology on over 50 physical-chemical properties or biologically active compounds.

MDF SPR/SAR methodology

The MDF SAR method integrates the complex information obtained from the structure of the compounds into models in order to explain the activity/property of interest. The input data for a given set of molecules are represented by the molecular and/or structural formulas and the experimental

values for the activity and/or property of interest. By applying the methodology, the molecular descriptors and/or uni- versus multivariate models are obtained.

The following six steps were used in the modelling process [32].

÷ Step 1: drawing of the topological model (2D) of the compounds for each molecule from the set of interest by using the HyperChem [33].
÷ Step 2: building the geometrical model (3D) of each compound from the set of interest by using HyperChem [33].
÷ Step 3: applying a semiempirical model (for calculating the partial charge distribution on atoms) and (sometimes) a quantum mechanics model (up to the most advanced ones such as *Ab-initio* and *Time-Dependent Density Functional Theory*) using specific modules of HyperChem [33] (examples: HyperNewton, HyperGauss, HyperNDO) in order to obtain an optimized geometrical model in vitro or in vivo.
÷ Step 4: generating the molecular descriptors family by using the MDF software. The MDF software is described below.
÷ Step 5: applying the bias procedure in order to eliminate identical descriptors from the generated family.
÷ Step 6: obtaining simple and/or multivariate linear regression relationships between the members of the descriptors and the given property/activity. The following criteria are used [34,35]: (1) the goodness-to-fit of the model [36,37] (the correlation coefficient and the squared correlation coefficient; values close to ±1 indicate a good model); (2) the co-linearity between pairs of descriptors (values lower than 0.5 indicate the absence of co-linearity between descriptors); and (3) the significance of the regression model (for a significance level of 5%). The internal validation of the MDF SAR models is analyzed in cross-validation leave-one-out analysis [38]. A correlated correlation analysis is applied whenever appropriate by using Steiger's Z test at a significance level of 5% [27].

MDF SAR physical model

The MDF SAR approach has a mathematical model that comprises seven pieces, every piece having a list of possibilities, which come from the physical approach. Every piece provides a letter in the descriptor's name:

÷ The linearizing operator (thefirst letter) makes the link between micro, nano, and macro levels. Example: pH = -log[H^+] its macro property (measure, effect) measured in micro environment (phenomenon, cause), the presence and the number of H^+ in a given solution. It takes six values:

I (identity), *i* (inverse), *A* (absolute), *a* (inverse of absolute), *L* (logarithm of absolute), *l* (logarithm).

÷ The molecular level superposing operator (the second letter) lay upon the fragmental contributions. Its existence is sustained by the variety of the causality of the molecular property/activity, from specificity, regioselectivity, and selectivity (which most biological activities have) to the independent structural formula (such as relative mass - same for all molecular formula isomers). It takes nineteen values: *sized group* (*m* = smallest; *M* = largest; *n* = smallest absolute; *N* = largest absolute); *averaged group* (*S* = sum; *A* = average over all values; *a* = sum divided by the number of all fragments; *B* = average first by atom group and then by the whole molecule; *b* = adjusted *B* by bonds); *geometric group* (*P* = multiplication; *G* = geometric mean; *g* = adjusted *G* by fragments; *F* = geometric mean first by atom group and then by the whole molecule; *f* = adjusted *F* by bonds); *harmonic group* (*s* = harmonic sum, *H* = harmonic mean, *h* = adjusted *H* by fragments, *I* = harmonic mean first by atom group and then by the whole molecule, and *i* = adjusted *I* by bonds).

÷ The pair-based fragmentation criteria (the third letter) implements different criteria. Some parts of a molecule are more active and give the most of the activity/property of a molecule than others (substituent's role) as it was observed in the first SAR studies carried out by Hammet. It takes four values: *m* (minimal), *M* (maximal), *D* (Szeged, distance based), and *P* (Cluj, shortest paths based [22,39]).

÷ The interaction model (the fourth letter) implements different levels of approximation (scalar and vectorial) for superposing the descriptors of interaction at the fragment level. It is well known that a series of field-type interactions (such as gravitational and electrostatic) are vectorially treated at low range and scalarly treated at distance. It takes six values: *scalar* (*R* = rare model and resultant relative to fragment's head, *r* = rare model and resultant relative to conventional origin, *M* = medium model and resultant relative to fragment's head, *m* = medium model and resultant relative to conventional origin), *vectorial* (*D* = dense model and resultant relative to fragment's head, *d* = dense model and resultant relative to conventional origin).

÷ The interaction descriptor (the fifth letter) implements a series of interaction descriptors for physical entities (such as force, field, energy, potential), as they occur in magnetism, electrostatics, gravity and quantum mechanics. Different physical entities have different formulas. It takes twenty-four values: *D* (distance), *d* (inverted distance), *O* (first atom's property), *o* (inverted of first atom's property), *P* (product of atomic

property), *p* (inverted *P*), Q (squared *P*), *q* (inverted *Q*), *J* (first atom's property multiplied by distance), *j* (inverted *J*), *K* (product of atomic property and distance), *k* (inverted *K*), *L* (product of distance and squared atomic property), *l* (inverted *L*), *V* (first atom's property potential), *E* (first atom's property field), *W* (first atom's property work), *w* (property's work), *F* (first atom's property's force), *f* (property's force), *S* (first atom's property's weak nuclear force), *s* (property's weak nuclear force), *T* (first atom's property strong nuclear force), t (property's strong nuclear force).

÷ The stomic property (the sixth letter) discriminates atoms through elemental properties. Every atom has a series of characteristics and/or properties making it similar and/or dissimilar to another. It takes six values: *M* (mass), *Q* (charge), *C* (cardinality), *E* (electronegativity), *G* (group electronegativity), *and H* (number of attached atoms of hydrogen).

÷ The distance operator (the seventh letter) implements both 2D and 3D approaches (topology and geometry). It takes two values: *g* (geometry), *t* (topology).

The application of every piece of mathematical model is a physical model, each model being able to take more than one value.

The molecular descriptors family is generated following the calculation of 787968 (6 × 19 × 4 × 6 × 24 × 6 × 2) possibilities. Not all these possibilities have physical meaning (e.g. the logarithm of a negative number). Moreover, not all of them produce finite numbers (e.g. the division by zero). For a given set of molecules a descriptor can be degenerated relative to the set (having the same value for all the molecules in the set) and relatively to another descriptor (two descriptors with different calculation formulas produce the same result for all molecules in the set). A bias procedure trails out these descriptors from the family of the set. Depending on the set, the number of MDF members is around 100000.

MDF software

The MDF software was created by using the triad: PHP (Pre Hypertext Processor, [40]), MySQL database [41], and FreeBSD server [42]. A set of programs completes the task of generating the molecular descriptors family.

Two databases (one temporary `MDFSARtmp`- for the sets being processed and one permanent `MDFSARs` - for finalized sets) stored on a FreeBSD server from IntraNet [IP 172.27.211.5] by using a MySQL database server. On December 25, 2007, `MDFSARtmp` had 174 tables (1.8 Gb), and `MDFSARs` more than 300 tables (> 3.5Gb). Four tables were generated for each investigated set:

÷ `"NameSet"_tmpx` (787968/6 = 131328 records; fields: molecules; records: descriptors)
÷ `"NameSet"_data` (field: property/activity; records: molecules; number of records equal with number of molecules)
÷ `"NameSet"_valx` (fields: molecules; records: descriptors; after bias has about 100000 records)
÷ `"NameSet"_valy` (records: same number of as `"NumeSet"_valx` table; fields: M(X), M(X*X), M(X*Y), r^2(X,Y), Name(X), where M = average operator, r^2 = determination coefficient, Name = name of X, Y = property/activity, X = MDF member).

Note that the numeric fields of the `"NameSet"_valy` table are computed for multivariate regression purposes (significantly decreasing the execution time). The `0_MDFSARRes` table (one per database) contains all obtained MDF SARs (fields: name (of the set), eq (MDF SAR), r^2 (determination coefficient), m (molecule's number), n (number of MDF members in MDF SAR model).

A set of five PHP applications generated MDF, running on a FreeBSD server from IntraNet [IP 172.27.211.4]:

÷ *0_mdf_prepare.php*: creates the structure for the set tables using the name of the directory (for set name) and names of the files (for molecules names)
÷ *1_mdf_generate.php*: generates MDF for the set of interest, filling 131328 records in the `"NameSet"_tmpx` table for every molecule. It is a multitasking application, one task being executed for every molecule at the same time.
÷ *2_mdf_linearize.php*: it applies the linearizing operator (131328 × 6 = 787968) and it fills valid records (having sense and finite) into the `"NameSet"_xval` and the `"NameSet"_yval` tables.
÷ *3_mdf_bias.php*: it sorts them in its memory by r^2 and it deletes the degenerations from both the `"NameSet"_xval` and the `"NameSet"_yval` tables simultaneously.
÷ *4_mdf_order.php*: it sorts them in its memory by r^2 again, it creates two temporary tables containing ordered records by r^2 from `"NameSet"_xval` and `"NameSet"_yval`, and it deletes the old tables and renames the new ones.

Our previous experience in working with a great number of molecular descriptors [22] indicated that the best found descriptor (the one which correlates the best with the measured property) is never found among the descriptors of the best found structure-activity relationship with two

descriptors. Thus, MDF uses pairs of descriptors when we search for SAR models with more than one descriptor (natural selection).

MDF uses a genetic algorithm for QSPR/QSAR modelling (genetic algorithms are a particular class of evolutionary algorithms, being categorized as global search heuristics [43]). The peculiarities of the genetic algorithm used are:

÷ Step 1 (inheritance and mutation). The linearization procedure described above is applied to the solution domain (2 × 6 × 24 × 6 × 4 × 19 MDF members having a genetic representation with six letters) whenever every descendent is obtained from a parent (inheritance) through a transformation (mutation). The number of descendants obtained is six times higher than that of the parents. In this step, the fitness function is defined as having real and distinct values. Over half of the descendants die due to mutation (around 300000 descendants remain, which now have genetic representations with seven letters).

÷ Step 2 (selection). A bias procedure (selection) is applied to the solution domain (MDF descendants from Step 1). In this step, the fitness function is defined as having distinct first nine digits of the determination coefficient. Only around 100000 members pass the selection process. Another selection is made from this solution domain obtained: the best descriptor (the one that best correlates with the measured property).

÷ Step 3 (crossover). Pairs of MDF members are crossed over in order to obtain models with two descriptors. Two fitness functions are used here: “to have the best determination coefficient” and “to have the best cross-validation leave-one-out score”.

Searching procedures of the uni- and multivariate models were created using Delphi client-server programs [44].

MDF structure- Property/activity relationships discussed

More than one hundred and forty models are presented. Seventy properties or activities on different classs of compounds (thirty-one) have been investigated. The results are structured on two sections (properties - MDF SPR Models and activities - MDF SAR models) based on the investigated property/activity and the class of compounds.

The univariate regression model was presented for the investigated class of compounds. The model with two descriptors was also presented whenever possible [45].

The statistics associated with the MDF model are expressed as correlation coefficient (r), standard error of estimated (s); Fisher parameter of the model and associated type I error (F(p)) at a significance level of 5%. The statistics obtained in cross-validation leave-one-out analysis are expressed as correlation coefficient ($r_{cv\text{-}loo}$), standard error of predicted ($s_{cv\text{-}loo}$), Fisher parameter of the predictive model and associated type I error ($F_{cv\text{-}loo}(p)$) [34] at a significance level of 5%.

MDF SPR models

In order to justify the introduction of the molecular descriptors family on the structure-property relationships a series of twenty properties on nine classes of compounds have been investigated. The results obtained are presented and discussed below.

Partition & activity coefficients

Volatile organic compounds - n-octanol/water partition coefficient

The analysis of the results reveals that the model with two descriptors obtained better performances in terms of goodness-of-fit and cross-validation [34]. According to the model with two descriptors, the octanol/water partition coefficient of the investigated volatile organic compounds is of geometrical and topological nature and it depends on the partial charge and number of the directly bounded hydrogens.

Sample size [reference]	24 [46]	
MDF SPR Equation	$\hat{y}=-0.004\cdot x+2.09$	$\hat{y}=0.65\cdot x_1-0.13\cdot x_2+3.99$
SPR Determination (%)	53	81
MDF Descriptor(s): x_1 & x_2	ISDRTHg	LsPrDQt & IADRSHg
Dominant Atomic Property	Hydrogen's (H)	Charge (Q) & Hydrogen (H)
Interaction Via	Space (geometry)	Bonds (topology) & Space (geometry)
Interaction Model	$H^2\cdot d^{-4}$	d & $H^2\cdot d^{-3}$
Structure on Property Scale	Identity	Logarithmic & Identity
Model Statistics	r=0.7309; s=0.43;F(p)= 25 ($4.98\cdot10^{-5}$)	r=0.9006; s=0.28; F (p)=45 ($2.51\cdot10^{-8}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.6792; $s_{cv\text{-}loo}$=0.47; $F_{cv\text{-}loo}$ (p)=18 ($3.00\cdot10^{-4}$)	$r_{cv\text{-}loo}$=0.8815; $s_{cv\text{-}loo}$=0.31; $F_{cv\text{-}loo}$ (p)=36.41 ($1.49\cdot10^{-7}$)

Para-substituted phenols - n-octanol/water partition coefficient

The octanol/water partition coefficient of para substituted phenols depended directly on the molecules' geometry and it was related with partial charge and group electronegativity (see the model with two descriptors).

Sample size [reference]	30 [47]	
MDF SPR Equation [reference]	$\hat{y}=-923.42\cdot x+4.25$	$\hat{y}=0.003\cdot x_1-0.40\cdot x_2+1.07$ [48]
SPR Determination (%)	71	89
MDF Descriptor(s): x_1 & x_2	IsPdOQg	isDDkGg & IMmrKQg
Dominant Atomic Property	Charge (Q)	Group Electronegativity (G) & Charge (Q)
Interaction Via	Space (geometry)	Space (geometry) & Space (geometry)
Interaction Model	Q	$G^{-2}\cdot d^{-1}$ & $Q^{-2}\cdot d^{-1}$
Structure on Property Scale	Identity	Inversed & Identity
Model Statistics	r=0.8412; s=0.60; F (p) = 68 ($5.83\cdot10^{-9}$)	r=0.9457; s=0.37; F (p) = 114 ($6.65\cdot10^{-14}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8115; $s_{cv\text{-}loo}$=0.65; $F_{cv\text{-}loo}$ (p)=53 ($5.82\cdot10^{-8}$)	$r_{cv\text{-}loo}$=0.9306; $s_{cv\text{-}loo}$=0.41; $F_{cv\text{-}loo}$ (p)=87 ($1.69\cdot10^{-12}$)

Polychlorinated biphenyls - n-octanol/water partition coefficient

According to the model with two descriptors, the octanol/water partition coefficient of investigated polychlorinated biphenyls is of geometrical nature and depends on the atomic and group electronegativity.

Sample size [reference]	206 [49]	
MDF SPR Equation	$\hat{y}=2552.7\cdot x-14.22$	$\hat{y}=-0.44\cdot x_1+0.04\cdot x_2+3.12$
SPR Determination (%)	87	89
MDF Descriptor(s): x_1 & x_2	iBMmwHg	IIDDKGg & IHDRKEg
Dominant Atomic Property	Hydrogen (H)	Group Electronegativity (G) & Electronegativity (E)
Interaction Via	Space (geometry)	Space (geometry) & Space (geometry)
Interaction Model	$H^2\cdot d^{-1}$	$G^2\cdot d$ & $E^2\cdot d$
Structure on Property Scale	Inversed	Identity & Identity
Model Statistics	r=0.9347; s=0.30; F (p)=1410 ($7.75\cdot10^{-94}$)	r=0.9433; s= 0.28; F (p)= 819 ($1.71\cdot10^{-98}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9330; $s_{cv\text{-}loo}$=0.30; $F_{cv\text{-}loo}$ (p)=1372 ($9.13\cdot10^{-93}$)	$r_{cv\text{-}loo}$=0.9409; $s_{cv\text{-}loo}$=0.28; $F_{cv\text{-}loo}$ (p)=784 ($9.33\cdot10^{-97}$)

Organic pollutants - Soil/water partition coefficient normalized to organic carbon

The soil/water partition coefficient of the studied organic pollutants proved to be a topological property related with the number of directly bounded hydrogen atoms and atomic electronegativity (see the equation with two descriptors).

Sample size [reference]	8 [50]	
MDF SPR Equation	$\hat{y}=-17.45\cdot x+8.12$	$\hat{y}=-0.22\cdot x_1-0.68\cdot x_2+16.62$
SPR Determination (%)	90	98
MDF Descriptor(s): x_1 & x_2	IbPMtMt	lfDMWHt & IbmrTEt
Dominant Atomic Property	Mass (M)	Hydrogen (H) & Electronegativity (E)
Interaction Via	Bonds (topology)	Bonds (topology) & Bonds (topology)
Interaction Model	$M^2\cdot d^{-4}$	$H^2\cdot d^{-1}$ & $E^2\cdot d^{-4}$
Structure on Property Scale	Identity	Logarithmic & Identity
Model Statistics	r=0.9483; s=0.21; F (p)=54 ($3.33\cdot10^{-4}$)	r=0.9839; s=0.22; F (p)=26 ($2.27\cdot10^{-3}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8710; $s_{cv\text{-}loo}$=0.34; $F_{cv\text{-}loo}$ (p)=18 ($5.27\cdot10^{-3}$)	$r_{cv\text{-}loo}$=0.9481; $s_{cv\text{-}loo}$=0.24; $F_{cv\text{-}loo}$ (p)=22 ($3.30\cdot10^{-3}$)

Fifteen standard amino acids - Partition (1st & 2nd columns) & activity coefficients (3rd column)

Fifteen standard amino acids were investigated: alanine (Ala), asparagine (Asn), aspartate (Asp), cysteine (Cys), glutamine (Gln), glutamate (Glu), glycine (Gly), isoleucine (Ile), leucine (Leu), lysine (Lys), methionine (Met), phenylalanine (Phe), serine (Ser), threonine (Thr), and valine (Val).

Sample size [reference]	15 [33]	15 [51]	15 [33]
MDF SPR Equation [reference]	$\hat{y}=-1.37\cdot x-0.87$ [52]	$\hat{y}=-4.87\cdot x+6.42$ [52]	$\hat{y}=-44.59\cdot x+18.09$ [52]
SPR Determination (%)	90	93	93
MDF Descriptor(s): x_1 & x_2	lGDdKQg	lHMrqQg	iGMmLQt
Dominant Atomic Property	Charge (Q)	Charge (Q)	Charge (Q)
Interaction Via	Space (geometry)	Space (geometry)	Bonds (topology)
Interaction Model	$Q^{-2}\cdot d^{-1}$	$Q^{-1}(\sqrt{Q})^{-1}$	$Q\cdot d$
Structure on Property Scale	Logarithmic	Logarithmic	Inversed
Model Statistics	r=0.9472; s=0.31; F (p)=113 ($8.61\cdot10^{-8}$)	r=0.9658; s=0.34; F (p)= 180 ($5.34\cdot10^{-9}$)	r=0.9620; s=0.87; F (p)= 161 ($1.06\cdot10^{-8}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9303; $s_{cv\text{-}loo}$=0.35; $F_{cv\text{-}loo}$ (p)=83 ($5.12\cdot10^{-7}$)	$r_{cv\text{-}loo}$=0.9536; $s_{cv\text{-}loo}$=0.40; $F_{cv\text{-}loo}$ (p)= 130 ($3.82\cdot10^{-8}$)	$r_{cv\text{-}loo}$=0.9440; $s_{cv\text{-}loo}$=1.06; $F_{cv\text{-}loo}$ (p)=106 ($1.30\cdot10^{-7}$)

The partition coefficients of the 15 amino acids were studied on two different scales. In both situations (column 1 and 2) the property was identified as being of geometrical nature and also depended on the partial charge. The activity coefficient is of topological nature but it also depends on the partial charge.

Chromatographic parameters

Polychlorinated biphenyls - Relative retention time

The relative retention time of polychlorinated biphenyls proved to be of both topological and geometrical nature, and was related with the number of directly bounded hydrogens (see the model with two descriptors).

Sample size [reference]	206 [49]	
MDF SPR Equation	$\hat{y}=0.09\cdot x-0.17$	$\hat{y}=0.02\cdot x_1-1.02\cdot x_2-5.99$
SPR Determination (%)	98	99.7
MDF Descriptor(s): x_1 & x_2	iIDRwHg	ISDmsHt & lADrtHg
Dominant Atomic Property	Hydrogen (H)	Hydrogen (H) & Hydrogen (H)
Interaction Via	Space (geometry)	Bonds (topology) & Space (geometry)
Interaction Model	$H^2\cdot d^{-1}$	$H^2\cdot d^{-3}$ & $H^2\cdot d^{-4}$
Structure on Property Scale	Inversed	Identity & Logarithmic
Model Statistics	r=0.9921; s=0.02; F (p)=13013 ($1.64\cdot10^{-189}$)	r=0.9986; s=0.01; F (p)=36600 ($1.10\cdot10^{-265}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9920; $s_{cv\text{-}loo}$=0.02; $F_{cv\text{-}loo}$ (p)=12777 ($1.06\cdot10^{-188}$)	$r_{cv\text{-}loo}$=0.9985; $s_{cv\text{-}loo}$=0.01; $F_{cv\text{-}loo}$ (p)=35416 ($3.33\cdot10^{-264}$)

Polychlorinated biphenyls - Relative response factor

The MDF SPR abilities to estimate and predict the relative response factor are not strong, the SAR determination being lower than 70%. According to the model with two descriptors, the relative response factor of the polychlorinated biphenyls is both of topological and geometrical nature and it strongly depends on the number of directly bounded hydrogens.

Sample size [reference]	209 [49]	
MDF SPR Equation	$\hat{y}=0.53\cdot x-0.51$	$\hat{y}=-357.3\cdot x_1+2.16\cdot x_2+5.08$
SPR Determination (%)	63	69
MDF Descriptor(s): x_1 & x_2	iHMdTHg	imMrFHt & iHDdFHg
Dominant Atomic Property	Hydrogen (H)	Hydrogen (H) & Hydrogen (H)

Table. Continued

Interaction Via	Space (geometry)	Bonds (topology) & Space (geometry)
Interaction Model	$H^2 \cdot d^{-4}$	$H^2 \cdot d^{-2}$ & $H^2 \cdot d^{-2}$
Structure on Property Scale	Inversed	Inversed & Inversed
Model Statistics	r=0.7929; s=0.22; F (p)=351 ($1.67 \cdot 10^{-46}$)	r=0.8324; s=0.20; F (p)=232 ($9.59 \cdot 10^{-54}$
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.7873; $s_{cv\text{-}loo}$=0.22; $F_{cv\text{-}loo}$ (p)=337 ($2.02 \cdot 10^{-45}$)	$r_{cv\text{-}loo}$=0.8258; $s_{cv\text{-}loo}$=0.20; $F_{cv\text{-}loo}$ (p)=221 ($3.73 \cdot 10^{-52}$)

Organophosphorus herbicides - Retention chromatography index

The retention chromatography index of organophosphorus herbicides proved to be estimated and predicted by using the MDF approach. The SPR determination was higher than 90%. According to the model with two descriptors the retention chromatography index is of topological and geometrical nature and depends on relative atomic mass and partial charge.

Sample size [reference]	10 [53]	
MDF SPR Equation [reference]	$\hat{y}=0.32 \cdot x-3.37$	$\hat{y}=6.37 \cdot x_1+0.06 \cdot x_2-62.36$ [23]
SPR Determination (%)	94	99.92
MDF Descriptor(s): x_1 & x_2	IBPdqHg	lSDmwMt & iHPDEQg
Dominant Atomic Property	Hydrogen (H)	Mass (M) & Charge (Q)
Interaction Via	Space (geometry)	Bonds (topology) & Space (geometry)
Interaction Model	$1/H\sqrt{H}$	$M^2 \cdot d^{-1}$ & $M \cdot d^{-2}$
Structure on Property Scale	Logarithmic	Logarithmic & Inversed
Model Statistics	r=0.9708; s=0.78; F (p)=131 ($3.08 \cdot 10^{-6}$)	r=0.9996; s=0.10; F (p) = 4348 ($1.47 \cdot 10^{-11}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9563; $s_{cv\text{-}loo}$=0.95; $F_{cv\text{-}loo}$ (p)=85 ($1.55 \cdot 10^{-5}$)	$r_{cv\text{-}loo}$=0.9993; $s_{cv\text{-}loo}$=0.13; $F_{cv\text{-}loo}$ (p)=2344 ($1.28 \cdot 10^{-10}$)

Molar refraction

Cyclic organophosphorus (1st & 2nd column) & fifteen standard amino acids (3rd column)

The MDF SPR approach proved to be a very good approach. The determination coefficient is higher than or equal with 98 percent. The molar refraction proved to be of topological nature and related with the relative atomic mass and the atomic electronegativity (see the model with two variables obtained for the cyclic organophosphorus s). In a SPR determination of 98% the molar refraction proved to be of geometrical nature, linearly related with the paretial charge in the sample of standard amino acids.

Sample size [reference]	10 [53]		15 [51]
MDF SPR Equation [reference]	$\hat{y}=16.37\cdot x-0.28$	$\hat{y}=28.25\cdot x_1-83.97\cdot x_2+17.39$ [54]	$\hat{y}=-0.89\cdot x+6.7$ [52]
SPR Determination (%)	99	100	98
MDF Descriptor(s): x_1 & x_2	iIMdsGg	lGDmSMt & lAmrfEt	lFMMwQg
Dominant Atomic Property	Group electronegativity (G)	Mass (M) & Electronegativity (E)	Charge (Q)
Interaction Via	Space (geometry)	Bonds (topology) & Bonds (topology)	Space (geometry)
Interaction Model	$G^2\cdot d^{-3}$	$M^2\cdot d^{-3}$ & $E^2\cdot d^{-2}$	$Q^2\cdot d^{-1}$
Structure on Property Scale	Inversed	Logarithmic & Logarithmic	Logarithmic
Model Statistics	r=0.9959; s=0.96; F (p)=975 ($1.20\cdot10^{-9}$)	r=0.9999; s=0.07; F (p)=83206 ($4.44\cdot10^{-16}$)	r=0.9892; s=1.13; F (p)=590 ($3.22\cdot10^{-12}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9942; $s_{cv\text{-}loo}$=1.15; $F_{cv\text{-}loo}$ (p)=679 ($5.04\cdot10^{-9}$)	$r_{cv\text{-}loo}$=0.9999; $s_{cv\text{-}loo}$=0.11; $F_{cv\text{-}loo}$ (p)=40592 ($5.99\cdot10^{-15}$)	$r_{cv\text{-}loo}$=0.9845; $s_{cv\text{-}loo}$=1.35; $F_{cv\text{-}loo}$ (p)= 409 ($3.29\cdot10^{-11}$)

Boiling point

Alkanes

Good determinations are obtained in the estimation and prediction of the boiling point of the investigated alkanes. The best performing model, the one with two descriptors, revealed that the boiling point of the investigated compound is of topological nature and directly related with group electronegativity and the number of directly bounded hydrogens (see the model with two descriptors).

Sample size [reference]	73 [55]	
MDF SPR Equation	$\hat{y}=188.40\cdot x-507.95$	$\hat{y}=-67.45\cdot x_1+4.89\cdot x_2-129.20$
SPR Determination (%)	99	99.8
MDF Descriptor(s): x_1 & x_2	lbMdsHg	lGDrtGt & IbDrfHt
Dominant Atomic Property	Hydrogen (H)	Group Electronegativity (G) & Hydrogen (H)
Interaction Via	Space (geometry)	Bonds (topology) & Bonds (topology)
Interaction Model	$H^2\cdot d^{-3}$	$G^2\cdot d^{-4}$ & $H^2\cdot d^{-2}$
Structure on Property Scale	Logarithmic	Inversed & Identity
Model Statistics	r=0.9956; s=3.81; F (p)=8050 ($1.23\cdot10^{-75}$)	r=0.9991; s=1.75; F (p)=19361 ($4.66\cdot10^{-99}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9954; $s_{cv\text{-}loo}$=3.91; $F_{cv\text{-}loo}$ (p)=7654 ($7.44\cdot10^{-75}$)	$r_{cv\text{-}loo}$=0.9990; $s_{cv\text{-}loo}$=1.82; $F_{cv\text{-}loo}$ (p)=17837 ($8.87\cdot10^{-98}$)

Other properties

Fifteen standard amino acids - Magnetic susceptibility (1st column) & dipole moment (2nd column) & solubility (3rd column)

The magnetic susceptibility of the investigated standard amino acids revealed to be of geometrical nature and linearly dependent on partial charge. The dipole moment and the solubility of the studied standard amino acids are of topological nature and depend on partial charge. For the last two properties (dipole moment and solubility) the model had the same molecular descriptor, but the performance was better for solubility than for the dipole moment.

Sample size [reference]	15 [51]	15 [51]	15 [51]
MDF SPR Equation [reference]	$\hat{y}$=-92.99·x+84.21 [52]	$\hat{y}$=-8.70·x+0.19 [52]	$\hat{y}$=-25.28·x+4.06 [52]
SPR Determination (%)	91	79	87
MDF Descriptor(s): x_1 & x_2	iHMRqQg	IiDRLQt	IiDRLQt
Dominant Atomic Property	Charge (Q)	Charge (Q)	Charge (Q)
Interaction Via	Space (geometry)	Bonds (topology)	Bonds (topology)
Interaction Model	$Q^{-1}(\sqrt{Q})^{-1}$	d·Q	Q·d
Structure on Property Scale	Inversed	Identity	Identity
Model Statistics	r=0.9548; s=2.98; F (p)=134 ($3.20\cdot10^{-8}$)	r=0.8885; s=0.50; F (p)=49 ($9.61\cdot10^{-6}$)	r=0.9338; s=1.08; F (p)=89 ($3.63\cdot10^{-7}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9381; $s_{cv\text{-}loo}$=3.48; $F_{cv\text{-}loo}$ (p)=95 ($2.42\cdot10^{-7}$)	$r_{cv\text{-}loo}$=0.8339; $s_{cv\text{-}loo}$=0.61; $F_{cv\text{-}loo}$ (p)= 29 ($1.18\cdot10^{-4}$)	$r_{cv\text{-}loo}$=0.9070; $s_{cv\text{-}loo}$=1.27; $F_{cv\text{-}loo}$ (p) = 60 ($3.15\cdot10^{-6}$)

Fifteen standard amino acids - Hückel energy (1st column) & hydration energy (2nd column)

The Hückel energy of the investigated amino acids revealed to be a geometrical property that depends on atomic electronegativity. Hydration energy revealed to be of topological nature and related with partial charge. Note that the determination coefficient is higher than 90% in both models.

Sample size [reference]	15 [33]	15 [33]
MDF SPR Equation [referenc	$\hat{y}$=868.02·x-1417.5 [52]	$\hat{y}$=17.59·x+19.46 [52]
SPR Determination (%)	99.7	93
MDF Descriptor(s): x_1 & x_2	lfPdkEg	iGPmLQt

Table. Continued

Dominant Atomic Property	Electronegativity (E)	Charge (Q)
Interaction Via	Space (geometry)	Bonds (topology)
Interaction Model	$E^{-2} \cdot d^{-1}$	$Q \cdot d$
Structure on Property Scale	Logarithmic	Inversed
Model Statistics	r=0.9983; s=235; F (p)=3915 ($1.53 \cdot 10^{-18}$)	r=0.9646; s=0.86; F (p)=174 ($6.74 \cdot 10^{-9}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9979; $s_{cv\text{-}loo}$=264; $F_{cv\text{-}loo}$ (p)=3089 ($1.11 \cdot 10^{-16}$)	$r_{cv\text{-}loo}$=0.9503; $s_{cv\text{-}loo}$=1.01; $F_{cv\text{-}loo}$ (p)=121 ($5.90 \cdot 10^{-8}$)

Fifteen standard amino acids - Polarizability (1st column) & refractivity (2nd column)

The polarizability and refractivity of the investigated standard amino acids was estimated by the same molecular descriptor. Thus, the properties are of geometrical nature and directly related with atomic electronegativity.

Sample size [reference]	15 [33]	15 [33]
MDF SPR Equation [referenc	$\hat{y}$=36.97·x-4.84 [52]	$\hat{y}$=93.72·x-13.09 [52]
SPR Determination (%)	98	97
MDF Descriptor(s): x_1 & x_2	iIMdWEg	iIMdWEg
Dominant Atomic Property	Electronegativity (E)	Electronegativity (E)
Interaction Via	Space (geometry)	Space (geometry)
Interaction Model	$E^{2} \cdot d^{-1}$	$E^{2} \cdot d^{-1}$
Structure on Property Scale	Inversed	Inversed
Model Statistics	r=0.9883; s=0.46; F (p)=546 ($5.32 \cdot 10^{-12}$)	r=0.9862; s=1.27; F (p)=462 ($1.53 \cdot 10^{-11}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9825; $s_{cv\text{-}loo}$=0.56; $F_{cv\text{-}loo}$ (p)=362 ($7.10 \cdot 10^{-11}$)	$r_{cv\text{-}loo}$=0.9794; $s_{cv\text{-}loo}$=1.55; $F_{cv\text{-}loo}$ (p)=306 ($2.03 \cdot 10^{-10}$)

MDF-SAR models

The abilities of the molecular descriptors family approach have been investigated on twenty-one samples of biological active compounds. A total number of seventy-three activities were investigated on different classes of compounds.

Water activated carbon adsorption

Organic compounds

The water activated carbon adsorption on the investigated organic compounds revealed to be of topological and geometrical nature and related with the number of directly bounded hydrogens and partial charge (see the

Table. Continued

Sample size [reference]	16 [56]	
MDF SAR Equation [reference]	$\hat{y}$=-57.99·x+1.99 [57]	$\hat{y}$=0.85·x_1+0.003·x_2+2.58 [57]
SAR Determination (%)	86	98
MDF Descriptor(s): x_1 & x_2	iSDrDQt	IiMMWHt & lPMDVQg
Dominant Atomic Property	Charge (Q)	Hydrogen (H) & Charge (Q)
Interaction Via	Bonds (topology)	Bonds (topology) & Space (geometry)
Interaction Model	d	$H^2 \cdot d^{-1}$ & $Q \cdot d^{-1}$
Structure on Activity Scale	Inversed	Identity & Logarithmic
Model Statistics	r=0.9270; s=0.13; F (p)=86 ($2.43 \cdot 10^{-7}$)	r=0.9905; s=0.05; F (p)=337 ($6.30 \cdot 10^{-12}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8959; $s_{cv\text{-}loo}$=0.16; $F_{cv\text{-}loo}$ (p)=56 ($2.92 \cdot 10^{-6}$)	$r_{cv\text{-}loo}$=0.9873; $s_{cv\text{-}loo}$=0.06; $F_{cv\text{-}loo}$ (p)=251 ($4.14 \cdot 10^{-11}$)

model with two descriptors). The power of determination of the activity is good - close to optimum (98%).

Hydrophobic vs. hydrophilic character

The hydrophobic or hydrophilic character, which is an important property in protein structure and protein-protein interactions, is one of the most studied properties of the amino acids. Many hydrophobicity scales have already been reported (see below). The differences between scales are significant: Janin (1979) and Kyte & Doolittle (1982) classified cysteine as the most hydrophobic while Wolfenden et al. [58] or Rose et al. [59] did not. These differences could be explained by the fundamentally different methods used for constructing the scale.

Fifteen standard amino acids – Hydrophobicity

The hydrophobicity on the Bumble, Hessa et al. and Kyte & Doolittle scales revealed to be of geometrical nature and directly related with partial charge on the sample of fifteen standard amino acids (see the table below). Excepting the Bumble scale, the MDF models obtained had a determination coefficient higher than 90%.

Sample size [reference]	15 [51]	15 [60]	15 [61]
MDF SAR Equation [reference]	$\hat{y}$=-160·x-0.07 [52]	$\hat{y}$=8.5·x-0.58 [52]	$\hat{y}$=-21·x+12 [52]

Table. Continued

SAR Determination (%)	65	90.5	95
MDF Descriptor(s)	AbmrEQg	iMDRoQg	IGDROQg
Dominant Atomic Property	Charge (Q)	Charge (Q)	Charge (Q)
Interaction Via	Space (geometry)	Space (geometry)	Space (geometry)
Interaction Model	$Q \cdot d^2$	Q^{-1}	Q
Structure on Activity Scale	Proportional	Inversed	Proportional
Model Statistics	r=0.8085; s=1.68; F (p)=25 ($2.64 \cdot 10^{-4}$)	r=0.9514; s=0.44; F (p)=124 ($5.05 \cdot 10^{-8}$)	r=0.9759; s=0.71; F (p)=260 ($5.66 \cdot 10^{-10}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.7550; $s_{cv\text{-}loo}$=1.88; $F_{cv\text{-}loo}$ (p)=17 ($1.21 \cdot 10^{-3}$)	$r_{cv\text{-}loo}$=0.9351; $s_{cv\text{-}loo}$=0.51; $F_{cv\text{-}loo}$ (p)=90 ($3.26 \cdot 10^{-7}$)	$r_{cv\text{-}loo}$=0.9659; $s_{cv\text{-}loo}$=0.80; $F_{cv\text{-}loo}$ (p)=203 ($2.57 \cdot 10^{-9}$)

Twenty standard amino acids – Hydrophobicity

The sample of twenty standard amino acids used for the following MDF SAR models contains: alanine (Ala), arginine (Arg), asparagine (Asn), aspartate (Asp), cysteine (Cys), glutamine (Gln), glutamate (Glu), glycine (Gly), histidine (His), isoleucine (Ile), leucine (Leu), lysine (Lys), methionine (Met), phenylalanine (Phe), proline (Pro), serine (Ser), threonine (Thr), tryptophan (Trp), tyrosine (Tyr), and valine (Val).

Sample size [reference]	20 [62]	20 [63]	20 [64]
MDF SAR Equation [reference]	$\hat{y}$=0.39·x-1.23 [65]	$\hat{y}$=-27.79·x+6.55 [65]	$\hat{y}$=-1.73·x-2.88 [65]
SAR Determination (%)	44	66	69
MDF Descriptor(s): x_1 & x_2	amMRLQt	immRoQg	LmDROQg
Dominant Atomic Property	Charge (Q)	Charge (Q)	Charge (Q)
Interaction Via	Bonds (topology)	Space (geometry)	Space (geometry)
Interaction Model	Q·d	Q^{-1}	Q
Structure on Activity Scale	Inversed	Inversed	Logarithmic
Model Statistics	r=0.6649; s=1.21; F (p)=14 ($1.38 \cdot 10^{-3}$)	r=0.8163; s=2.19; F (p)=6 ($1.14 \cdot 10^{-5}$)	r=0.8309; s=1.70; F (p)=40 ($5.70 \cdot 10^{-6}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.5961; $s_{cv\text{-}loo}$=1.37; $F_{cv\text{-}loo}$ (p)=7 ($1.44 \cdot 10^{-2}$)	$r_{cv\text{-}loo}$=0.7740; $s_{cv\text{-}loo}$=2.41; $F_{cv\text{-}loo}$ (p) = 27 ($6.50 \cdot 10^{-5}$)	$r_{cv\text{-}loo}$=0.7936; $s_{cv\text{-}loo}$= 1.87; $F_{cv\text{-}loo}$ (p) = 30 ($3.34 \cdot 10^{-5}$)

Two out of three of the above presented models for hydrophobicity proved that the activity was of geometrical nature and depended on the partial charge. None of the above models was strong; the coefficient of determination was lower than 70%.

Twenty standard amino acids – Hydrophobicity

The MDF SAR model for hydrophobicity on the Wimley & White scale is of topological nature and depends on partial charge. The hydrophobicity models on the Hoop & Woods and Cowan & Whittaker scales revealed to be of geometrical nature and depended on partial charge. The coefficient of determination associated with the above presented models proved not to be powerful, even if the values were higher than 50%.

Sample size [reference]	20 [66]	20 [67]	20 [68]
MDF SAR Equation [reference]	$\hat{y}$=7.35·x-3.37 [65]	$\hat{y}$=10.63·x-1.99 [65]	$\hat{y}$=-6.57·x+1.47 [65]
SAR Determination (%)	71	74	75
MDF Descriptor	iBmrWQt	iMPRoQg	AmDROQg
Dominant Atomic Property	Charge (Q)	Charge (Q)	Charge (Q)
Interaction Via	Bonds (topology)	Space (geometry)	Space (geometry)
Interaction Model	Q^2/d	Q^{-1}	Q
Structure on Activity Scale	Inversed	Inversed	Absolute
Model Statistics	r=0.8434; s=0.48; F (p)=44 ($3.00\cdot10^{-6}$)	r=0.8608; s=1.01; F (p)=52 ($1.11\cdot10^{-6}$)	r=0.8661; s=0.6; F (p)=54 ($1.15\cdot10^{-8}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.800; $s_{cv\text{-}loo}$=0.54; $F_{cv\text{-}loo}$ (p) = 32 ($2.25\cdot10^{-5}$)	$r_{cv\text{-}loo}$=0.8288; $_{loo}$=1.11; $F_{cv\text{-}loo}$ (p) = 39 ($6.49\cdot10^{-6}$)	$r_{cv\text{-}loo}$=0.8344; $_{loo}$=0.73; $F_{cv\text{-}loo}$ (p) = 41 ($7.94\cdot10^{-8}$)

Twenty standard amino acids – Hydrophobicity

Hydrophobicity on the Manavalan & Ponnuswamy, the Fauchere et al., and the Rao & Argos scales proved to be of geometrical nature and linearly

Sample size [reference]	20 [69]	20 [70]	20 [71]
MDF SAR Equation [reference]	$\hat{y}$=23.43·x+14.55 [65]	$\hat{y}$=5.94·x-4.36 [65]	$\hat{y}$=-2.73·x+1.43 [65]
SAR Determination (%)	78	78	79

Table. Continued

MDF Descriptor	inMrpQg	ibDRPQg	AmDROQg
Dominant Atomic Property	Charge (Q)	Charge (Q)	Charge (Q)
Interaction Via	Space (geometry)	Space (geometry)	Space (geometry)
Interaction Model	Q^{-2}	Q^{2}	Q
Structure on Activity Scale	Inversed	Inversed	Proportional
Model Statistics	r=0.8814; s=0.76; F (p)=63 ($2.84\cdot10^{-7}$)	r=0.8832; s=0.50; F (p)=65 ($2.50\cdot10^{-7}$)	r=0.8901; s=0.24; F (p)=69 ($1.48\cdot10^{-7}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8546; $s_{cv\text{-}loo}$=0.84; $F_{cv\text{-}loo}$ (p)=49 ($1.65\cdot10^{-6}$)	$r_{cv\text{-}loo}$=0.8611; $s_{cv\text{-}loo}$=0.54; $F_{cv\text{-}loo}$ (p) = 51 ($1.13\cdot10^{-6}$)	$r_{cv\text{-}loo}$=0.8545; $s_{cv\text{-}loo}$=0.28; $F_{cv\text{-}loo}$ (p) = 48 ($1.78\cdot10^{-6}$)

depended on partial charge. All the above MDF SAR models presented a weak coefficient of determination (the values were higher than 75%). At this value, there is an important linear relationship between the molecular descriptor and the hydrophobic or hydrophilic character.

Twenty standard amino acids – Hydrophobicity

All three models revealed that the hydrophobic or hydrophilic character was of geometrical nature and depended on partial charge. The determination was fairly good in these models (greater than 80%).

Sample size [reference]	20 [72]	20 [73]	20 [59]
MDF SAR Equation [reference]	$\hat{y}$=1.74·x+0.86 [65]	$\hat{y}$=-3.78·x+5.30 [65]	$\hat{y}$=1.75·x+0.86 [65]
SAR Determination (%)	81	85	90
MDF Descriptor	inMrpQg	IAmrLQg	inMrpQg
Dominant Atomic Property	Charge (Q)	Charge (Q)	Charge (Q)
Interaction Via	Space (geometry)	Space (geometry)	Space (geometry)
Interaction Model	Q^{-2}	Q·d	$Q^{2}\cdot d^{-3}$
Structure on Activity Scale	Inversed	Identity	Inversed
Model Statistics	r=0.8974; s=0.05; F (p)=76 ($6.76\cdot10^{-8}$)	r=0.9208; s =0.80; F (p)=100 ($8.69\cdot10^{-9}$)	r=0.8974; s=0.05; F (p)=74 ($8.21\cdot10^{-8}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8744; $s_{cv\text{-}loo}$=0.06; $F_{cv\text{-}loo}$ (p) = 56 ($6.37\cdot10^{-7}$)	$r_{cv\text{-}loo}$=0.9073; $s_{cv\text{-}loo}$=0.68; $F_{cv\text{-}loo}$ (p) = 84 ($3.48\cdot10^{-8}$)	$r_{cv\text{-}loo}$=0.8744; $s_{cv\text{-}loo}$=0.06; $F_{cv\text{-}loo}$ (p) = 58 ($4.73\cdot10^{-7}$)

Twenty standard amino acids – Hydrophobicity

The hydrophobicity on the Urry scale proved to be of topological nature and linearly dependent on the atomic electronegativity of the standard amino acids studied. The hydrophobicity on the Engelman et al., and on the Eisenberg et al. scales proved to be of geometrical nature and directly dependent on the partial charge. The determination was considered rather good.

Sample size [reference]	20 [74]	20 [75]	20 [76]
MDF SAR Equation [reference]	$\hat{y}$=-11.96·x-29.73 [65]	$\hat{y}$=-753.09·x+ 1.85 [65]	$\hat{y}$=-0.92·x+ 1.68 [65]
SAR Determination (%)	82	83	83
MDF Descriptor	iBDMkEt	INPrWQg	IAMdKQg
Dominant Atomic Property	Electronegativity (E)	Charge (Q)	Charge (Q)
Interaction Via	Bonds (topology)	Space (geometry)	Space (geometry)
Interaction Model	$Q^{-2}\cdot d^{-1}$	$Q^{2}\cdot d^{-1}$	$Q^{2}\cdot d$
Structure on Activity Scale	Inversed	Logarithmic	Logarithmic
Model Statistics	r=0.9047; s=1.07; F (p)=81 ($4.40\cdot10^{-8}$)	r=0.9116; s=2.07; F (p)=89 ($2.26\cdot10^{-8}$)	r=0.9128; s=0.42; F (p)=90 ($2.02\cdot10^{-8}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8819; $s_{cv\text{-}loo}$=1.18; $F_{cv\text{-}loo}$ (p) = 63 ($2.85\cdot10^{-7}$)	$r_{cv\text{-}loo}$=0.8731; $s_{cv\text{-}loo}$=2.56; $F_{cv\text{-}loo}$ (p) = 51 ($1.13\cdot10^{-6}$)	$r_{cv\text{-}loo}$=0.8935; $s_{cv\text{-}loo}$=0.46; $F_{cv\text{-}loo}$ (p) = 70 ($1.31\cdot10^{-7}$)

Twenty standard amino acids – Hydrophobicity

The hydrophobicity on the Cowan et al. scale is of geometrical nature and depends on charge. The same is valid for the model with fifteen amino acids. This observation is also true for the hydrophobicity on the Hessa et al.

Sample size [reference]	20 [68]	20 [77]	20 [60]
MDF SAR Equation [reference]	$\hat{y}$=-2.16·x+4.64 [65]	$\hat{y}$=817.95·x+81.72 [65]	$\hat{y}$=7.18·x-0.41 [65]
SAR Determination (%)	84	85	85
MDF Descriptor(s)	lbmdKQg	inMrpQg	AmDROQg
Dominant Atomic Property	Charge (Q)	Charge (Q)	Charge (Q)
Interaction Via	Space (geometry)	Space (geometry)	Space (geometry)

Table. Continued

Interaction Model	$Q^2 \cdot d$	Q^{-2}	Q
Structure on Activity Scale	Logarithmic	Inversed	Proportional
Model Statistics	r=0.9182; s=0.52; F (p)=97 ($1.15 \cdot 10^{-8}$)	r=0.9232; s=20.73; F (p)=104 ($6.69 \cdot 10^{-9}$)	r=0.9238; s=0.32; F (p)=105 ($6.24 \cdot 10^{-9}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8984; $_{loo}$=0.58; $F_{cv\text{-}loo}$ (p) = 75 ($7.94 \cdot 10^{-8}$)	$r_{cv\text{-}loo}$=0.9082; $s_{cv\text{-}loo}$=22.58; $F_{cv\text{-}loo}$ (p) = 85 ($3.16 \cdot 10^{-8}$)	$r_{cv\text{-}loo}$=0.9018; $s_{cv\text{-}loo}$=0.58; $F_{cv\text{-}loo}$ (p) = 78 ($6.01 \cdot 10^{-8}$)

scale, the determination being lower than in the model obtained on the sample of fifteen amino acids. All models revealed that the hydrophobic or hydrophilic character was of geometrical nature and depended on the partial charge.

Twenty standard amino acids – Hydrophobicity

Hydrophobicity is of geometrical nature and depends on charge. In these models, the determination is slightly better compared with the above models in the sample of twenty standard amino acids.

Sample size [reference]	20 [78]	20 [79]	20 [61]
MDF SAR Equation [reference]	$\hat{y}$=-0.20·x+1.36 [65]	$\hat{y}$=1.85·x+11.05 [65]	$\hat{y}$=19.17·x-7.60 [65]
SAR Determination (%)	85	86	87
MDF Descriptor(s)	iIPmLQt	lfPROQg	IiDRLQt
Dominant Atomic Property	Charge (Q)	Charge (Q)	Charge (Q)
Interaction Via	Bonds (topology)	Space (geometry)	Bonds (topology)
Interaction Model	Q·d	Q	d·Q
Structure on Activity/ Property Scale	Inversed	Logarithmic	Identity
Model Statistics	r=0.9252; s=0.36; F (p)= 107 ($5.30 \cdot 10^{-9}$)	r=0.9259; s=2.46; F (p)=108 ($4.88 \cdot 10^{-9}$)	r=0.9328; s=1.11; F (p)=120 ($2.10 \cdot 10^{-9}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9003; $s_{cv\text{-}loo}$=0.42; $F_{cv\text{-}loo}$ (p) = 75 ($8.02 \cdot 10^{-8}$)	$r_{cv\text{-}loo}$=0.8935; $s_{cv\text{-}loo}$=2.97; $F_{cv\text{-}loo}$ (p) = 69 ($4.91 \cdot 10^{-8}$)	$r_{cv\text{-}loo}$=0.9226; $s_{cv\text{-}loo}$=1.18; $F_{cv\text{-}loo}$ (p) = 103 ($7.25 \cdot 10^{-9}$)

Twenty standard amino acids – Hydrophobicity

The best determination on the sample of twenty amino acids was obtained for the Black et al. scale. On the Monera et al. scale the determination was better but the sample was of nineteen amino acids (Proline was the amino acid excluded from the generation of molecular descriptors). As an overall conclusion, the hydrophobicity of amino acids is of geometrical nature and depends on partial charge.

Sample size [reference]	20 [80]	19 [81] (- Proline)
MDF SAR Equation [reference]	$\hat{y}$=-0.96·x+0.86 [65]	$\hat{y}$=843.88·x+86.05 [65]
SAR Determination (%)	88	90
MDF Descriptor(s)	lAmrLQg	inMrpQg
Dominant Atomic Property	Charge (Q)	Charge (Q)
Interaction Via	Space (geometry)	Space (geometry)
Interaction Model	$d\cdot\sqrt{Q}$	Q^{-2}
Structure on Activity Scale	Proportional	Inversed
Model Statistics	r=0.9376; s=0.12; F (p)=131 ($1.09\cdot10^{-9}$)	r=0.9504; s=16.49; F (p)=159 ($4.77\cdot10^{-10}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9263; $s_{cv\text{-}loo}$=0.13; $F_{cv\text{-}loo}$ (p)=109 ($4.73\cdot10^{-9}$)	$r_{cv\text{-}loo}$=0.9380; $s_{cv\text{-}loo}$=18.37; $F_{cv\text{-}loo}$ (p)=125 ($3.00\cdot10^{-9}$)

Toxicity

Polychlorinated organic compounds

The toxicity of the investigated polychlorinated organic compounds revealed to be of geometrical nature and related with partial charge on the univariate as well as on the MDF model with two descriptors.

Sample size [reference]	31 [82]	
MDF SAR Equation	$\hat{y}$=-9.06·x+4.00	$\hat{y}$=-8.33·x_1+0.28·x_2+0.83
SAR Determination (%)	72	87
MDF Descriptor(s): x_1 & x_2	IsMRKQg	IsMRKQg & AHPROQg
Dominant Atomic Property	Charge (Q)	Charge (Q) & Charge (Q)
Interaction Via	Space (geometry)	Space (geometry) & Space (geometry)
Interaction Model	$Q^{-2}\cdot d^{-1}$	$Q^{-2}\cdot d^{-1}$ & Q

Table. Continued

Structure on Activity Scale	Identity	Identity & Absolute
Model Statistics	r=0.8514; s =0.30; F (p)=76 ($1.27\cdot10^{-9}$)	r=0.9318; s=0.21; F (p)=92 ($4.79\cdot10^{-13}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8353; $s_{cv\text{-}loo}$=0.32; $F_{cv\text{-}loo}$ (p)=67 ($5.24\cdot10^{-9}$)	$r_{cv\text{-}loo}$=0.9118; $s_{cv\text{-}loo}$=0.24; $F_{cv\text{-}loo}$ (p)=68 ($1.68\cdot10^{-11}$)

Mono-substituted nitrobenzenes

The toxicity of the investigated mono-substituted nitrobenzenes is of both geometrical and topological nature. It also depends on group electronegativity and partial charge. Ninety-six percent of variation in toxicity could be explained by the linear relationship with two molecular descriptors.

Sample size [reference]	39 [83]	
MDF SAR Equation	$\hat{y}=-91.15\cdot x+6.27$	$\hat{y}=-92.37\cdot x_1-7.28\cdot x_2+6.37$
SAR Determination (%)	60	60
MDF Descriptor(s): x_1 & x_2	IBMrkGg	IBMrkGg & IsPmVQt
Dominant Atomic Property	Group Electronegativity (G)	Group Electronegativity (G) & Charge (Q)
Interaction Via	Space (geometry)	Space (geometry) & Bonds (topology)
Interaction Model	$G^{-2}\cdot d^{-1}$	$E^{-2}\cdot d^{-1}$ & $Q\cdot d^{-1}$
Structure on Activity Scale	Identity	Identity & Identity
Model Statistics	r=0.7717; s=0.35; F (p)=54 ($8.87\cdot10^{-9}$)	r=0.7739; s=0.35; F (p)=27 ($7.33\cdot10^{-8}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.7474; $s_{cv\text{-}loo}$=0.37; $F_{cv\text{-}loo}$ (p)=48 ($4.71\cdot10^{-8}$)	$r_{cv\text{-}loo}$=0.6947; $s_{cv\text{-}loo}$=0.41; $F_{cv\text{-}loo}$ (p)=16 ($1.25\cdot10^{-5}$)

Benzene derivates

The toxicity of the investigated benzene derivates revealed to be of topological and geometrical nature. It also depended on partial charge and on the number of directly bounded hydrogens.

Sample size [reference]	69 [84]	
MDF SAR/SPR Equation	$\hat{y}=-0.91\cdot x+2.92$	$\hat{y}=-9.66\cdot x_1+1.00\cdot x_2+3.25$
SAR/SPR Determination (%)	68	87
MDF Descriptor(s): x_1 & x_2	lFPdoGg	ABmrsQg & iGPrfHt

Table. Continued

Dominant Atomic Property	Group Electronegativity (G)	Charge (Q) & Hydrogen (H)
Interaction Via	Space (geometry)	Space (geometry) & Bonds (topology)
Interaction Model	G^{-1}	$Q^2 \cdot d^{-3}$ & $H^2 \cdot d^{-2}$
Structure on Activity/ Property Scale	Logarithmic	Absolute & Inversed
Model Statistics	r=0.8262; s=0.43; F (p)=144 ($1.87 \cdot 10^{-18}$)	r=0.9331; s=0.28; F (p)=222 ($1.48 \cdot 10^{-30}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8160; $s_{cv\text{-}loo}$=0.44; $F_{cv\text{-}loo}$ (p)=133 ($1.08 \cdot 10^{-17}$)	$r_{cv\text{-}loo}$=0.9267; $s_{cv\text{-}loo}$=0.29; $F_{cv\text{-}loo}$ (p)=201 ($2.97 \cdot 10^{-29}$)

Alkyl metal compounds

The toxicity of alkyl metal compounds revealed to be of geometrical nature and depended on the relative atomic mass as well as on the partial charge. The MDF model was very good, its determination coefficient being close to 1.

Sample size [reference]	10 [85]	
MDF SAR Equation [reference]	$\hat{y}=0.25 \cdot x+0.33$	$\hat{y}=28.06 \cdot x_1+0.08 \cdot x_2+2.80$ [86]
SAR Determination (%)	97	99.7
MDF Descriptor(s): x_1 & x_2	iFDmdCg	IbMmpMg & LPPROQg
Dominant Atomic Property	Cardinality (C)	Mass (M) & Charge (Q)
Interaction Via	Space (geometry)	Space (geometry) & Space (geometry)
Interaction Model	d^{-1}	M^{-2} & Q
Structure on Activity Scale	Inversed	Identity & Logarithmic
Model Statistics	r=0.9830; s=0.19; F (p)=230 ($3.54 \cdot 10^{-7}$)	r=0.9988; s=0.06; F (p)=1473 ($6.49 \cdot 10^{-10}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9729; $s_{cv\text{-}loo}$=0.24; $F_{cv\text{-}loo}$ (p)=141 ($2.29 \cdot 10^{-6}$)	$r_{cv\text{-}loo}$=0.9980; $s_{cv\text{-}loo}$=0.07; $F_{cv\text{-}loo}$ (p)=841 ($4.57 \cdot 10^{-9}$)

Para-substituted phenols – Toxicity

The model obtained for the toxicity of the para substituted phenols revealed to be a good model. This model showed that the toxicity was of topological and geometrical nature and was related with partial charge.

Sample size [reference]	30 [47]	
MDF SAR Equation	$\hat{y}=-603.71\cdot x+1.74$	$\hat{y}=0.04\cdot x_1-0.22\cdot x_2-2.26$
SAR Determination (%)	71	90
MDF Descriptor(s): x_1 & x_2	IsPdOQg	ASMmVQt & lfDdOQg
Dominant Atomic Property	Charge (Q)	Charge (Q) & Charge (Q)
Interaction Via	Space (geometry)	Bonds (topology) & Space (Geometry)
Interaction Model	Q	$Q\cdot d^{-1}$ & Q
Structure on Activity Scale	Identity	Absolute & Logarithmic
Model Statistics	r=0.8458; s=0.38; F (p)=70 ($4.01\cdot10^{-9}$)	r=0.9472; s=0.23; F (p)=118 ($4.56\cdot10^{-14}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8219; $s_{cv\text{-}loo}$=0.41; $F_{cv\text{-}loo}$ (p)=58 ($2.68\cdot10^{-8}$)	$r_{cv\text{-}loo}$=0.9352; $s_{cv\text{-}loo}$=0.26; $F_{cv\text{-}loo}$ (p)=93 ($7.57\cdot10^{-13}$)

Para-substituted phenols - Relative toxicity

The relative toxicity of the investigated para-substituted phenols revealed to be of geometrical nature and linearly dependant on the partial charge and number of directly bounded hydrogens. The model with two descriptors could be considered considerably good, its determination coefficient being of eighty-five percent.

Sample size [reference]	30 [47]	
MDF SAR Equation	$\hat{y}=4.50\cdot x+6.59$	$\hat{y}=-1.76x_1-1.42\cdot x_2+12.14$
SAR Determination (%)	68	85
MDF Descriptor(s): x_1 & x_2	InDDoQg	AHMMVQg & inDmwHg
Dominant Atomic Property	Charge (Q)	Charge (Q) & Hydrogen (H)
Interaction Via	Space (geometry)	Space (geometry) & Space (geometry)
Interaction Model	Q^{-1}	$Q\cdot d^{-1}$ & $Q^2\cdot d^{-1}$
Structure on Activity Scale	Identity	Absolute & Inversed
Model Statistics	r=0.8275; s=0.54; F (p)=61 ($1.71\cdot10^{-8}$)	r=0.9225; s=0.38; F (p)=77 ($6.87\cdot10^{-12}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8011; $s_{cv\text{-}loo}$=0.57; $F_{cv\text{-}loo}$ (p)=50 ($1.07\cdot10^{-7}$)	$r_{cv\text{-}loo}$=0.9054; $s_{cv\text{-}loo}$=0.42; $F_{cv\text{-}loo}$ (p)=61 ($9.98\cdot10^{-11}$)

Quinoline – Cytotoxicity

The model with two descriptors obtained in the investigation of cytotoxicity in the quinoline sample showed that the activity was of topological

nature and depended on relative atomic mass and partial charge. The model was successfully able to predict the activity (the determination coefficient was close to 1). The prediction power was sustained by the value of the cross-validation leave-one-out score, which had a value of 0.9805 (see the model with two descriptors).

Sample size [reference]	15 [87]	
MDF SAR Equation [reference]	$\hat{y}=-6.58\cdot x+4.63$	$\hat{y}=8.35\cdot x_1+1.96\cdot x_2-4.49$ [88]
SAR Determination (%)	65	98
MDF Descriptor(s): x_1 & x_2	iHPMdCg	INDRLQt & lHPmTMt
Dominant Atomic Property	Cardinality (C)	Charge (Q) & Mass (M)
Interaction Via	Space (geometry)	Bonds (topology) & Bonds (topology)
Interaction Model	d^{-1}	$Q\cdot d$ & $M^2\cdot d^{-4}$
Structure on Activity Scale	Inversed	Identity & Inversed
Model Statistics	r=0.8044; s=0.63; F (p)=24 ($2.99\cdot10^{-4}$)	r=0.9882; s=0.17; F (p) = 250 ($1.65\cdot10^{-10}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.7576; $s_{cv\text{-}loo}$=0.70; $F_{cv\text{-}loo}$ (p)=17 ($1.21\cdot10^{-3}$)	$r_{cv\text{-}loo}$=0.9805; $s_{cv\text{-}loo}$=0.22; $F_{cv\text{-}loo}$ (p)=149 ($3.34\cdot10^{-9}$)

Quinoline – Mutagenicity

The model with two descriptors demonstrated that mutagenicity was of geometrical nature and strongly dependent on partial charge. The estimation power of the model was very close to 1.

Sample size [reference]	14 [87]	
MDF SAR Equation [reference]	$\hat{y}=0.008\cdot x-4.14$	$\hat{y}=0.21\cdot x_1+0.09\cdot x_2-1.57$ [88]
SAR Determination (%)	72	96
MDF Descriptor(s): x_1 & x_2	aAmrKQt	lNMrSQg & ASPrVQg
Dominant Atomic Property	Charge (Q)	Charge (Q) & Charge (Q)
Interaction Via	Bonds (topology)	Space (geometry) & Space (geometry)
Interaction Model	$Q^2\cdot d$	$Q^2\cdot d^{-3}$ & $Q\cdot d^{-1}$
Structure on Activity Scale	Inversed	Logarithmic & Proportional
Model Statistics	r=0.8456; s=0.44; F (p)=30 ($1.39\cdot10^{-4}$)	r=0.9782; s=0.18; F (p)=122 ($3.12\cdot10^{-8}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8054; $s_{cv\text{-}loo}$=0.49; $F_{cv\text{-}loo}$ (p)=21 ($6.46\cdot10^{-4}$)	$r_{cv\text{-}loo}$=0.9666; $s_{cv\text{-}loo}$=0.22; $F_{cv\text{-}loo}$ (p)=78 ($3.18\cdot10^{-7}$)

Insecticidal & herbicidal activities

Neonicotinoids - Insecticidal Activity

The model with two descriptors obtained a very good estimation power. According to this model the activity was of both geometrical and topological nature. It also depended directly on atomic electronegativity and partial charge.

Sample size [reference]	8 [89]	
MDF SAR Equation [reference]	$\hat{y}$=-77.49·x +19.15	$\hat{y}$=-2.21·x_1 +3.74·x_2 +43.34 [90]
SAR Determination (%)	88	99.9
MDF Descriptor(s): x_1 & x_2	iIDrSMg	ImMdsEg & lIMMFQt
Dominant Atomic Property	Mass (M)	Electronegativity (E) & Charge (Q)
Interaction Via	Space (geometry)	Space (geometry) & Bonds (topology)
Interaction Model	$M^2 \cdot d^{-3}$	$E^2 \cdot d^{-3}$ & $Q^2 \cdot d^{-2}$
Structure on Activity Scale	Inversed	Identity & Inversed
Model Statistics	r=0.9370; s=0.48; F (p)=43 ($5.95 \cdot 10^{-4}$)	r=0.9996; s=0.04; F (p)=2865 ($2.24 \cdot 10^{-8}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8692; $s_{cv\text{-}loo}$=0.68; $F_{cv\text{-}loo}$ (p)=18 ($5.49 \cdot 10^{-3}$)	$r_{cv\text{-}loo}$=0.9991; $s_{cv\text{-}loo}$=0.06; $F_{cv\text{-}loo}$ (p)=1386 ($1.37 \cdot 10^{-7}$)

Substituted triazines (Triazines) - Herbicidal activity

A very good MDF model for estimation and prediction was obtained for the herbicidal activity of substituted triazines. The herbicidal activity was of both geometrical and topological nature and depended on the number of directed bounded hydrogen and partial charge.

Sample size [reference]	30	
MDF SAR Equation [reference]	$\hat{y}$=-4284.7·x+7.47	$\hat{y}$=-8112.2·x_1+194.35·x_2+5.52
SAR Determination (%)	95	98
MDF Descriptor(s): x_1 & x_2	iSDRFHg	iSMMWHg & iSMmEQt
Dominant Atomic Property	Hydrogen (H)	Hydrogen (H) & Charge (Q)
Interaction Via	Space (geometry)	Space (geometry) & Bonds (topology)
Interaction Model	$H^2 \cdot d^2$	
Structure on Activity Scale	Inversed	Inversed & Inversed
Model Statistics	r=0.9754; s=0.16; F (p)=549 ($2.18 \cdot 10^{-20}$)	r=0.9876; s=0.13; F (p) = 533 ($1.37 \cdot 10^{-23}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9725; $s_{cv\text{-}loo}$=0.17; $F_{cv\text{-}loo}$ (p)=488 ($1.09 \cdot 10^{-19}$)	$r_{cv\text{-}loo}$=0.9855; $s_{cv\text{-}loo}$=0.12; $F_{cv\text{-}loo}$ (p)=449 ($1.52 \cdot 10^{-22}$)

Therapeutically activities

3-Indolyl derivates - Antioxidant efficacy

The antioxidant efficacy of the investigated 3-indolyl derivates revealed to be of geometrical nature and depended on partial charge (see the model with one descriptor). As was expected the model with two descriptors obtained better results (determination coefficient of 99.9%), but this model is close to the limit according to the Hawkins criteria [45].

Sample size [reference]	8 [93]	
MDF SAR Equation [reference]	$\hat{y}=-1.34\cdot10^{-5}\cdot x-3.76$	$\hat{y}=-1.10\cdot x_1-33.24\cdot x_2+7.18$ [94]
SAR Determination (%)	90	99.9
MDF Descriptor(s): x_1 & x_2	asMmtQg	lbPMkHg & iAPrVGt
Dominant Atomic Property	Charge (Q)	Hydrogen (H) & Group Electronegativity (G)
Interaction Via	Space (geometry)	Space (geometry) & Bonds (topology)
Interaction Model	$Q^2\cdot d^{-4}$	$H^{-2}\cdot d^{-1}$ & $G\cdot d^{-1}$
Structure on Activity Scale	Inversed	Logarithmic & Inversed
Model Statistics	r=0.9508; s=0.21; F (p) = 56 ($2.87\cdot10^{-4}$)	r=0.9999; s=0.01; F (p) = 12591 ($5.55\cdot10^{-10}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9119; $s_{cv\text{-}loo}$=0.29; $F_{cv\text{-}loo}$ (p)=29 ($1.64\cdot10^{-3}$)	$r_{cv\text{-}loo}$=0.9997; $s_{cv\text{-}loo}$=0.02; $F_{cv\text{-}loo}$ (p) = 3877 ($1.05\cdot10^{-8}$)

Substituted N 4-methoxyphenyl benzamides - Antiallergic activity

The model with two descriptors indicated that the antiallergic activity of the investigated substituted N 4-methoxyphenyl benzamides was of both geometrical and topological nature and depended on group electronegativity and relative atomic mass.

Sample size [reference]	23 [95]	
MDF SAR/SPR Equation	$\hat{y}=0.03\cdot x+0.21$	$\hat{y}=-6.8\cdot10^{-5}\cdot x_1-1.3\cdot10^{-6}\cdot x_2+0.18$
SAR/SPR Determination (%)	92	98
MDF Descriptor(s): x_1 & x_2	InPrdQg	IFDDpGg & ISDrFMt
Dominant Atomic Property	Charge (Q)	Group Electronegativity (G) & Mass (M)
Interaction Via	Space (geometry)	Space (geometry) & Bonds (topology)
Interaction Model	d^{-1}	G^2 & $M^2\cdot d^{-2}$
Structure on Activity/ Property Scale	Identity	Identity & Identity
Model Statistics	r=0.9603; s=0.33; F (p)=249 ($4.05\cdot10^{-13}$)	r=0.9920; s=0.15; F (p)=616 ($4.87\cdot10^{-20}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9513; $s_{cv\text{-}loo}$=0.36; $F_{cv\text{-}loo}$ (p)=199 ($3.46\cdot10^{-12}$)	$r_{cv\text{-}loo}$=0.9910; $s_{cv\text{-}loo}$=0.16; $F_{cv\text{-}loo}$ (p)=540 ($1.99\cdot10^{-19}$)

Polyhydroxyxanthones - Antituberculotic activity

The antituberculotic activity of the investigated polyhydroxyxanthones was well estimated and predicted by the model with two descriptors. Its estimated power was of 99.7%. The activity was of geometrical nature and depended on partial charge and group electronegativity.

Sample size [reference]	10 [96]	
MDF SAR Equation [reference]	$\hat{y}=-0.07\cdot x+9.74$	$\hat{y}=2.32\cdot x_1+19.34\cdot x_2-19.11$ [97]
SAR Determination (%)	82	99.7
MDF Descriptor(s): x_1 & x_2	isDDoHg	lHPDOQg & IsMRKGg
Dominant Atomic Property	Hydrogen (H)	Charge (Q) & Group Electronegativity (G)
Interaction Via	Space (geometry)	Space (geometry) & Space (geometry)
Interaction Model	H^{-1}	$Q \& G^{-2}\cdot d^{-1}$
Structure on Activity Scale	Inversed	Logarithmic & Identity
Model Statistics	r=0.9082; s=0.23; F (p)=38 ($2.78\cdot10^{-4}$)	r=0.9987; s=0.03; F (p)=1327 ($9.33\cdot10^{-10}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8346; $s_{cv\text{-}loo}$=0.31; $F_{cv\text{-}loo}$ (p)=18 ($2.99\cdot10^{-3}$)	$r_{cv\text{-}loo}$=0.9974; $s_{cv\text{-}loo}$=0.04; $F_{cv\text{-}loo}$ (p)=663 ($1.05\cdot10^{-8}$)

Taxoids - Growth inhibition activity

A model with the estimated power of 92% was obtained during the investigation of the taxoids. According to the model with two descriptors, the growth inhibition activity of the investigated taxoids was of geometrical nature and was related with the number of directly bounded hydrogens.

Sample size [reference]	34 [98]	
MDF SAR Equation [reference]	$\hat{y}=0.89\cdot x-8.23$	$\hat{y}=0.002\cdot x_1+77.22\cdot x_2-17.7$ [99]
SAR Determination (%)	83	92
MDF Descriptor(s): x_1 & x_2	IHDrFHt	isMdTHg & IiDrQHg
Dominant Atomic Property	Hydrogen (H)	Hydrogen (H) & Hydrogen (H)
Interaction Via	Bonds (topology)	Space (geometry) & Space (geometry)
Interaction Model	$H^2\cdot d^{-2}$	$H^2\cdot d^{-4}$ & $H\sqrt{H}$
Structure on Activity Scale	Identity	Inversed & Identity
Model Statistics	r=0.9108; s=0.51; F (p)=156 ($7.75\cdot10^{-14}$)	r=0.9583; s=0.36; F (p)=174 ($2.86\cdot10^{-18}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9006; $s_{cv\text{-}loo}$=0.54; $F_{cv\text{-}loo}$ (p)=137 ($4.08\cdot10^{-13}$)	$r_{cv\text{-}loo}$=0.9507; $s_{cv\text{-}loo}$=0.39; $F_{cv\text{-}loo}$ (p)=146 ($2.22\cdot10^{-16}$)

HEPTA and TIBO derivatives - anti-HIV-1 potencies

The anti-HIV-1 potencies of the investigated HEPTA and TIBO derivatives revealed to be of geometrical nature and related with atomic and group electronegativity. The estimated power of the model was modest, but increasing the number of molecular descriptors to five provide a significantly better model [101].

Sample size [reference]	57 [100]	
MDF SAR Equation	$\hat{y}$=-3776.8·x+8.29	$\hat{y}$=-19.43·x_1+11.07·x_2-4.30
SAR Determination (%)	61	78
MDF Descriptor(s): x_1 & x_2	imMDPMg	lIDrFEg & iMMsGg
Dominant Atomic Property	Mass (M)	Electronegativity (E) & Group Electronegativity (G)
Interaction Via	Space (geometry)	Space (geometry) & Space (geometry)
Interaction Model	M^2	$E^2 \cdot d^{-2}$ &$G^2 \cdot d^{-3}$
Structure on Activity Scale	Inversed	Logarithmic & Logarithmic
Model Statistics	r=0.7812; s=0.95; F (p)=86 ($7.58 \cdot 10^{-13}$)	r=0.8849; s=0.71; F (p)=97 ($5.77 \cdot 10^{-19}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.7636; $s_{cv\text{-}loo}$=0.98; $F_{cv\text{-}loo}$ (p)=76 ($5.03 \cdot 10^{-12}$)	$r_{cv\text{-}loo}$=0.8750; $s_{cv\text{-}loo}$=0.74; $F_{cv\text{-}loo}$ (p)=88 ($5.18 \cdot 10^{-18}$)

Substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides (40846_1) - Inhibition activity on carbonic anhydrase I

The inhibition activity on carbonic anhydrase I of the substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides revealed to be of geometrical nature and depended on partial charge and relative atomic mass according to the model with two descriptors. When the number of descriptors increased, a model whose estimated power was 10% higher than that of the model with two descriptors was obtained [103].

Sample size [reference]	40 [102]	
MDF SAR Equation	$\hat{y}$=-0.008·x+0.66	$\hat{y}$=0.11·x_1+3.10·10^{-3}·x_2+1.74
SAR Determination (%)	63	81
MDF Descriptor(s): x_1 & x_2	isMRdQt	inPRlQg & lPDMqMg
Dominant Atomic Property	Charge (Q)	Charge (Q) & Mass (M)
Interaction Via	Bonds (topology)	Space (geometry) & Space (geometry)
Interaction Model	d^{-1}	$Q^{-2} \cdot d^{-1}$ &M√M
Structure on Activity Scale	Inversed	Inversed & Logarithmic
Model Statistics	r=0.7927; s=0.33; F (p)=64 ($1.09 \cdot 10^{-9}$)	r=0.8975; s=0.24; F (p)=77 ($6.95 \cdot 10^{-14}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.7787; $s_{cv\text{-}loo}$=0.34; $F_{cv\text{-}loo}$ (p)=58 ($3.38 \cdot 10^{-9}$)	$r_{cv\text{-}loo}$=0.8882; $s_{cv\text{-}loo}$=0.25; $F_{cv\text{-}loo}$ (p)=69 ($3.37 \cdot 10^{-13}$)

Substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides (40846_2) - Inhibition activity on carbonic anhydrase II

The inhibition activity of the carbonic anhydrase II of the substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides was of geometrical nature and depended on cardinality and partial charge (see the above model with two descriptors). When the investigation continued, a more powerful model was obtained (a model with 4 descriptors whose estimated power was 11% higher than that of the model with two descriptors [104]).

Sample size [reference]	40 [102]	
MDF SAR Equation	$\hat{y}=-2.49\cdot10^{-3}\cdot x+1.43$	$\hat{y}=2.44\cdot x_1+0.09\cdot x_2-4.45$
SAR Determination (%)	55	79
MDF Descriptor(s): x_1 & x_2	lPmrSMg	imDdSCg & iiMrqQg
Dominant Atomic Property	Mass (M)	Cardinality (C) & Charge (Q)
Interaction Via	Space (geometry)	Space (geometry) & Space (geometry)
Interaction Model	$M^2\cdot d^{-3}$	$C^2\cdot d^{-3}$ & $Q\sqrt{Q}$
Structure on Activity Scale	Logarithmic	Inversed & Inversed
Model Statistics	r=0.7422; s=0.35; F (p)=47 ($4.21\cdot10^{-8}$)	r=0.8862; s=0.25; F (p)=68 ($4.36\cdot10^{-13}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.7187; $s_{cv\text{-}loo}$=0.37; $F_{cv\text{-}loo}$ (p)=41 ($1.80\cdot10^{-7}$)	$r_{cv\text{-}loo}$=0.8697; $s_{cv\text{-}loo}$=0.26; $F_{cv\text{-}loo}$ (p)=57 ($4.60\cdot10^{-12}$)

Substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides (40846_4) - Inhibition activity on carbonic anhydrase IV

The analysis of the model with two descriptors revealed that the inhibition activity on carbonic anhydrase IV of the substituted 1,3,4- thiadiazole- and 1,3,4-thiadiazoline-disulfonamides was of geometrical and topological nature and depended on partial charge. The estimated power of the model was

Sample size [reference]	40 [102]	
MDF SAR Equation	$\hat{y}=-0.006\cdot x+0.40$	$\hat{y}=0.11\cdot x_1+9.98\cdot10^{-9}\cdot x_2+0.80$
SAR Determination (%)	56	75
MDF Descriptor(s): x_1 & x_2	iAPmsQt	inPRlQg & iHMMTQt
Dominant Atomic Property	Charge (Q)	Charge (Q) & Charge (Q)
Interaction Via	Bonds (topology)	Space (geometry) & Bonds (topology)
Interaction Model	$Q^2\cdot d^{-3}$	$Q^{-1}\cdot d^{-1}$ & $Q^2\cdot d^{-4}$
Structure on Activity Scale	Inversed	Inversed & Inversed
Model Statistics	r=0.7455; s=0.36; F (p)=48 ($3.41\cdot10^{-8}$)	r=0.8672; s=0.27; F (p)=56 ($6.21\cdot10^{-12}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.7230; $s_{cv\text{-}loo}$=0.38; $F_{cv\text{-}loo}$ (p)=42 ($1.38\cdot10^{-7}$)	$r_{cv\text{-}loo}$=0.8536; $s_{cv\text{-}loo}$=0.29; $F_{cv\text{-}loo}$ (p)=0.49 ($3.54\cdot10^{-11}$)

slightly moderate, but the increase to four in the number of descriptors increased the estimated power with 16% [105].

Dipeptides - Inhibition activity

A model with moderate estimation power was obtained when the inhibition activity of dipeptides was investigated. The model with two descriptors showed that the activity was of both topological and geometrical nature and depended on relative atomic mass and number of directly bounded hydrogens.

Sample size [reference]	58 [106]	
MDF SAR Equation	$\hat{y}=2.26\cdot10^{-4}\cdot x+1.94$	$\hat{y}=-1.47\cdot x_1+0.12\cdot x_2+2.57$
SAR Determination (%)	75	85
MDF Descriptor(s): x_1 & x_2	ISMrSGg	ibDMFHt & ISPdlMg
Dominant Atomic Property	Group Electronegativity (G)	Hydrogen (H) & Mass (M)
Interaction Via	Space (geometry)	Bonds (topology) & Space (geometry)
Interaction Model	$G^2\cdot d^{-3}$	$H^2\cdot d^{-2}$ &$M^{-1}\cdot d^{-1}$
Structure on Activity Scale	Identity	Inversed & Identity
Model Statistics	r=0.8656; s=0.51; F (p)=167 ($1.32\cdot10^{-18}$)	r=0.9208; s=0.40; F (p)=153 ($1.15\cdot10^{-23}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8551; $s_{cv\text{-}loo}$=0.52; $F_{cv\text{-}loo}$ (p) = 152 ($9.73\cdot10^{-18}$)	$r_{cv\text{-}loo}$=0.9135; $s_{cv\text{-}loo}$=0.41; $F_{cv\text{-}loo}$ (p)=139 ($1.27\cdot10^{-22}$)

2,4-diamino-5-(substituted-benzyl)-pyrimidines - Inhibition activity

The inhibition activity of the investigated 2,4-diamino-5-(substituted-benzyl)-pyrimidines revealed to be of both geometrical and topological nature. It was also related with the number of directly bounded hydrogens (see the model with two descriptors).

Sample size [reference]	67 [107]	
MDF SAR Equation	$\hat{y}=-707.32\cdot x+10.83$	$\hat{y}=-4.90\cdot x_1+2.31\cdot x_2+3.26$
SAR Determination (%)	73	86
MDF Descriptor(s): x_1 & x_2	ibMrEMt	lImrKHt & lIMDWHg
Dominant Atomic Property	Mass (M)	Hydrogen (H) & Hydrogen (H)
Interaction Via	Bonds (topology)	Bonds (topology) & Space (geometry)
Interaction Model	$M\cdot d^{-2}$	$H^{-2}\cdot d^{-1}$ & $H^2\cdot d^{-1}$
Structure on Activity Scale	Inversed	Logarithmic & Logarithmic
Model Statistics	r=0.8551; s=0.32; F (p)=177 ($2.42\cdot10^{-20}$)	r=0.9268; s=0.23; F (p)=195 ($2.06\cdot10^{-28}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8459; $s_{cv\text{-}loo}$=0.33; $F_{cv\text{-}loo}$ (p)=163 ($1.62\cdot10^{-19}$)	$r_{cv\text{-}loo}$=0.9195; $s_{cv\text{-}loo}$=0.24; $F_{cv\text{-}loo}$ (p)=175 ($4.05\cdot10^{-27}$)

Peptide analogues - Inhibition activity

The model with two descriptors obtained when the inhibition activity of the peptide analogues was investigated showed that the activity was of geometrical nature and strongly related with the relative atomic mass.

Sample size [reference]	47 [108]	
MDF SAR Equation	$\hat{y}=3202.8\cdot x-16.51$	$\hat{y}=240.54\cdot x_1-0.10\cdot x_2+0.94$
SAR Determination (%)	81	88
MDF Descriptor(s): x_1 & x_2	IHmRpMg	IHMdpMg & IHMdOMg
Dominant Atomic Property	Mass (M)	Mass (M) & Mass (M)
Interaction Via	Space (geometry)	Space (geometry) & Space (geometry)
Interaction Model	$M^2\cdot d^{-3}$	$M^2\cdot d^{-3}$ & M
Structure on Activity Scale	Identity	Identity & Identity
Model Statistics	r=0.8999; s=0.28; F (p)=192 ($5.16\cdot10^{-18}$)	r=0.9400; s=0.22; F (p)=167 ($8.04\cdot10^{-22}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8900; $s_{cv\text{-}loo}$=0.29; $F_{cv\text{-}loo}$ (p)=171 ($1.11\cdot10^{-16}$)	$r_{cv\text{-}loo}$=0.9325; $s_{cv\text{-}loo}$=0.23; $F_{cv\text{-}loo}$ (p)=147 ($1.10\cdot10^{-20}$)

Lethal/effective concentration

Ordnance compounds - Fertilization of *Sea Urchin*

The lethal/effective concentration of the investigated ordnance compounds in the fertilization of the *Sea Urchin* revealed to be of topological nature. It also depended on group electronegativity (see the model with one descriptor). The estimated power did not increase with the increase in the number of descriptors used by the models. This sustained the ability of the model with one descriptors to estimate and predict.

Sample size [reference]	8 [109]	
MDF SAR Equation [reference]	$\hat{y}=75.98\cdot x+3937.6$	$\hat{y}=-4291.5\cdot x_1-24751\cdot x_2+82488$ [110]
SAR Determination (%)	99.9	99.9
MDF Descriptor(s): x_1 & x_2	ISPRwGt	iSDmtQg & lAMrFEt
Dominant Atomic Property	Group electronegativity (E)	Charge (Q) & Electronegativity (E)
Interaction Via	Bonds (topology)	Space (geometry) & Bonds (topology)
Interaction Model	$E^2\cdot d^{-1}$	$Q^2\cdot d^{-4}$ & $E^2\cdot d^{-2}$
Structure on Activity Scale	Identity	Inversed & Logarithmic
Model Statistics	r=0.9996; s=192; F (p)=7754 ($1.44\cdot10^{-10}$)	r=0.9999; s=13; F (p) = 823154 ($1.61\cdot10^{-14}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9990; $s_{cv\text{-}loo}$=328; $F_{cv\text{-}loo}$ (p)=2645 ($3.62\cdot10^{-9}$)	$r_{cv\text{-}loo}$=0.9999; $s_{cv\text{-}loo}$=18; $F_{cv\text{-}loo}$ (p)=418264 ($8.74\cdot10^{-14}$)

Ordnance Compounds - Embryological development (1st column) & germination (2nd column) of *Sea Urchin*

The lethal/effective concentration in the embryological development of the *Sea Urchin* revealed to be of topological nature and depended on partial charge. This model had a very good estimation power.

The investigation of the lethal/ effective concentration in the germination of the *Sea Urchin* revealed that it was of topological nature and related with partial charge. In this model the power of estimation was good, but it was not as high as for the previous model.

Sample size [reference]	5 [109]	7 [109]
MDF SAR Equation	$\hat{y}$=-0.37·x-0.16	$\hat{y}$=-1.09·x-7.09
SAR Determination (%)	99.9	93
MDF Descriptor	LIMmwQt	lNPmfQt
Dominant Atomic Property	Charge (Q)	Charge (Q)
Interaction Via	Bonds (topology)	Bonds (topology)
Interaction Model	$Q^2 \cdot d^{-1}$	$Q^2 \cdot d^{-2}$
Structure on Activity Scale	Logarithmic	Logarithmic
Model Statistics	r=0.9997; s=0.02; F (p)=5677 ($5.15 \cdot 10^{-6}$)	r=0.9650; s=0.35; F (p)=68 ($4.32 \cdot 10^{-4}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9992; $s_{cv\text{-}loo}$=0.05; $F_{cv\text{-}loo}$ (p)=1220 ($5.16 \cdot 10^{-5}$)	$r_{cv\text{-}loo}$=0.9197; $s_{cv\text{-}loo}$=0.52; $F_{cv\text{-}loo}$ (p)=27 ($3.43 \cdot 10^{-3}$)

Ordnance compounds - Zoospore germination of *Green Macroalgae*

The lethal/effective concentration of ordnance compounds in the zoospore germination of the *Green Macroalgae* seemed to be of geometrical nature and depended on relative atomic mass and partial charge (see the model with two descriptors).

Sample size [reference]	8 [109]	
MDF SAR Equation [reference]	$\hat{y}$=0.06·x-1.50	$\hat{y}$=-0.004·x_1+11.11·x_2+21.58 [110]
SAR Determination (%)	89	99.9
MDF Descriptor(s): x_1 & x_2	aIDmjQg	iHDRkMg & inMrPQg
Dominant Atomic Property	Charge (Q)	Mass (M) & Charge (Q)
Interaction Via	Space (geometry)	Space (geometry) & Space (geometry)
Interaction Model	$Q^{-1} \cdot d^{-1}$	$M^{-2} \cdot d^{-1}$ & Q^2
Structure on Activity Scale	Inversed	Inversed & Inversed
Model Statistics	r=0.9435; s=0.39; F (p)=49 ($4.32 \cdot 10^{-4}$)	r=0.9996; s=0.04; F (p)=2942 ($2.10 \cdot 10^{-8}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9129; $s_{cv\text{-}loo}$=0.50; $F_{cv\text{-}loo}$ (p)=27 ($1.94 \cdot 10^{-3}$)	$r_{cv\text{-}loo}$=0.9988; $s_{cv\text{-}loo}$=0.06; $F_{cv\text{-}loo}$ (p)=1042 ($2.80 \cdot 10^{-7}$)

Ordnance compounds - Germling length of *Green Macroalgae*

The lethal/effective concentration of ordnance compounds in the germling length of the *Green Macroalgae* revealed to be of geometrical nature and related with the number of directly bounded hydrogens and partial charge according to the model with two descriptors. The estimation and prediction power of this model was almost perfect.

Sample size [reference]	8 [109]	
MDF SAR Equation [reference]	$\hat{y}$=-1.88·x-6.13	$\hat{y}$=-10.09·x_1-1.39·x_2+6.95 [110]
SAR Determination (%)	89	99.9
MDF Descriptor(s): x_1 & x_2	LIDmjQg	iGDREHg & lnDDVQg
Dominant Atomic Property	Charge (Q)	Hydrogen (H) & Charge (Q)
Interaction Via	Space (geometry)	Space (geometry) & Space (geometry)
Interaction Model	$Q^{-1}\cdot d^{-1}$	$H\cdot d^{-2}$ &$Q\cdot d^{-1}$
Structure on Activity Scale	Logarithmic	Inversed & Logarithmic
Model Statistics	r=0.9445; s=0.35; F (p)=50 ($4.09\cdot10^{-4}$)	r=0.9999; s=0.02; F (p)=11350 ($7.20\cdot10^{-10}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8969; $s_{cv\text{-}loo}$=0.49; $F_{cv\text{-}loo}$ (p)=23 ($3.09\cdot10^{-3}$)	$r_{cv\text{-}loo}$=0.9992; $s_{cv\text{-}loo}$=0.06; $F_{cv\text{-}loo}$ (p)=1089 ($2.51\cdot10^{-7}$)

Ordnance compounds - Germling cell number of *Green Macroalgae*

Two models were obtained for the lethal/effective concentration of ordnance compounds in the germling cell number of the *Green Macroalgae*. According to the model with two descriptors, the activity was of geometrical nature and strongly depended on partial charge. The estimated power of the model with two descriptors was very high.

Sample size [reference]	8 [109]	
MDF SAR Equation [reference]	$\hat{y}$=-1.87·x-6.02	$\hat{y}$=382.96·x_1-5.15·x_2+5.97 [110]
SAR Determination (%)	88	99.9
MDF Descriptor(s): x_1 & x_2	LIDmjQg	AHDmtQg & inMDqQg
Dominant Atomic Property	Charge (Q)	Charge (Q) & Charge (Q)
Interaction Via	Space (geometry)	Space (geometry) & Space (geometry)
Interaction Model	$Q^{-1}\cdot d^{-1}$	$Q^{2}\cdot d^{-2}$ & $Q^{-1}(\sqrt{Q})^{-1}$
Structure on Activity Scale	Logarithmic	Absolute & Inversed
Model Statistics	r=0.9359; s=0.38; F (p)=42 ($6.28\cdot10^{-4}$)	r=0.9996; s=0.03; F (p)=3132 ($1.80\cdot10^{-8}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8907; $s_{cv\text{-}loo}$=0.50; $F_{cv\text{-}loo}$ (p)=22 ($3.46\cdot10^{-3}$)	$r_{cv\text{-}loo}$=0.9992; $s_{cv\text{-}loo}$=0.05; $F_{cv\text{-}loo}$ (p)=1545 ($1.05\cdot10^{-7}$)

Ordnance compounds - Survival and reproductive success of *Polychaete* (1st column) & *Juveniles* survival of *Opossum Shrimp* (2nd column)

The lethal/effective concentration of ordnance compounds expressed as the successful reproduction and survival of *Polychaete Juveniles* (1st column) revealed to be of geometrical nature and related with partial charge.

The lethal/effective concentration of ordnance compounds expressed as the survival of the *Opossum Shrimp* (2nd column) also revealed to be of geometrical nature and depended on partial charge.

Sample size [reference]	7 [109]	7 [109]
MDF SAR Equation	$\hat{y}$=-102.72·x-0.79	$\hat{y}$=-1.31·x+0.28
SAR Determination (%)	97	91
MDF Descriptor(s)	IAPmtQt	LHDmjQg
Dominant Atomic Property	Charge (Q)	Charge (Q)
Interaction Via	Bonds (topology)	Space (geometry)
Interaction Model	$Q^2 \cdot d^{-4}$	$Q^{-1} \cdot d^{-1}$
Structure on Activity Scale	Identity	Logarithmic
Model Statistics	r=0.9835; s=0.22; F (p)=148 ($6.65 \cdot 10^{-5}$)	r=0.9531; s=0.25; F (p)=50 ($8.93 \cdot 10^{-4}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9748; $s_{cv\text{-}loo}$=0.27; $F_{cv\text{-}loo}$ (p)=94 ($1.98 \cdot 10^{-4}$)	$r_{cv\text{-}loo}$=0.9171; $s_{cv\text{-}loo}$=0.35; $F_{cv\text{-}loo}$ (p)=24 ($4.53 \cdot 10^{-3}$)

Ordnance compounds - *Redfish* larvae survival

The lethal/effective concentration of ordnance compounds expressed as *redfish* larvae survival proved to be of topological nature and related with cardinality and partial charge.

Sample size [reference]	8 [109]	
MDF SAR Equation [reference]	$\hat{y}$=16.91·x-1.73	$\hat{y}$=-14.72·x_1-0.11·x_2+17 [110]
SAR Determination (%)	93	99.9
MDF Descriptor(s): x_1 & x_2	anDRJQt	iAMrECt & aAPmfQt
Dominant Atomic Property	Charge (Q)	Cardinality (C) & Charge (Q)
Interaction Via	Bonds (topology)	Bonds (topology) & Bonds (topology)
Interaction Model	Q·d	$C \cdot d^{-2}$ & $Q^2 \cdot d^{-2}$
Structure on Activity Scale	Inversed	Inversed & Inversed
Model Statistics	r=0.9655; s=0.32; F (p)=82 ($9.99 \cdot 10^{-5}$)	r=0.9998; s=0.03; F (p)=6295 ($3.14 \cdot 10^{-9}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9408; $s_{cv\text{-}loo}$=0.42; $F_{cv\text{-}loo}$ (p)=45 ($5.10 \cdot 10^{-4}$)	$r_{cv\text{-}loo}$=0.9996; $s_{cv\text{-}loo}$=0.04; $F_{cv\text{-}loo}$ (p)=2813 ($2.35 \cdot 10^{-8}$)

No observed effect concentration (NOEC)

Ordnance compounds - Fertilization (1st column) & embryological development (2nd column) & germination of *Sea Urchin* (3rd column)

The NOEC of the investigated ordnance compounds in the fertilization of the *Sea Urchin* revealed to be of geometrical nature and related with compounds' cardinality.

The NOEC of the investigated ordnance compounds in the embryological development of the *Sea Urchin* revealed to be of geometrical nature and related with partial charge.

The NOEC of the investigated ordnance compounds in the germination of the *Sea Urchin* revealed to be of topological nature and related with partial charge.

All the above mentioned models had good estimation and prediction power, the coefficient of determination being higher than or equal with 95%.

Sample size [reference]	7 [109]	7 [109]	6 [109]
MDF SAR Equation	$\hat{y}=-0.80\cdot x+3.93$	$\hat{y}=0.17\cdot x+1.42$	$\hat{y}=0.001\cdot x-1.27$
SAR Determination (%)	96	95	97
MDF Descriptor	imMrtCg	ASPmwQg	asmrfQt
Dominant Atomic Property	Cardinality (C)	Charge (Q)	Charge (Q)
Interaction Via	Space (geometry)	Space (geometry)	Bonds (topology)
Interaction Model	$C^2\cdot d^{-4}$	$Q^2\cdot d^{-1}$	$Q^2\cdot d^{-2}$
Structure on Activity Scale	Inversed	Logarithmic	Inversed
Model Statistics	r=0.9787; s=0.10; F (p)=114 ($1.25\cdot10^{-4}$)	r=0.9738; s=0.08; F (p)=92 ($2.09\cdot10^{-4}$)	r=0.9859; s=0.27; F (p)=139 ($2.97\cdot10^{-4}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9627; $s_{cv\text{-}loo}$=0.13; $F_{cv\text{-}loo}$ (p)=61 ($5.58\cdot10^{-4}$)	$r_{cv\text{-}loo}$=0.9540; $s_{cv\text{-}loo}$=0.10; $F_{cv\text{-}loo}$ (p)=50 ($8.80\cdot10^{-4}$)	$r_{cv\text{-}loo}$=0.9704; $s_{cv\text{-}loo}$=0.39; $F_{cv\text{-}loo}$ (p) = 64 ($1.34\cdot10^{-3}$)

Ordnance compounds - Germling length and cell number of *Green Macroalgae*

The NOEC of the investigated ordnance compounds in the germling length and cell number of the *Green Macroalgae* revealed to be of both topological and geometrical nature. It also depended on group electronegativity and relative atomic mass (the models with two descriptors). The model with two descriptors proved to have excellent estimation and prediction power.

Sample size [reference]	8 [109]	
MDF SAR Equation	$\hat{y}=0.06\cdot x-1.74$	$\hat{y}=24.40\cdot x_1-28.28\cdot x_2+252.94$
SAR Determination (%)	87	99.9
MDF Descriptor(s): x_1 & x_2	aIDmjQg	lMMRSGt & lsPmpMg
Dominant Atomic Property	Charge (Q)	Group Electronegativity (G) & Mass (M)
Interaction Via	Space (geometry)	Bonds (topology) & Space (geometry)
Interaction Model	$Q^{-1}\cdot d^{-1}$	$G^{2}\cdot d^{-3}$ & M^{-2}
Structure on Activity Scale	Inversed	Logarithmic & Logarithmic
Model Statistics	r=0.9355; s=0.40; F (p)=42 ($6.38\cdot10^{-4}$)	r=0.9995; s=0.04; F (p)=2499 ($3.16\cdot10^{-8}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9003; $s_{cv\text{-}loo}$=0.53; $F_{cv\text{-}loo}$ (p)=22 ($3.17\cdot10^{-3}$)	$r_{cv\text{-}loo}$=0.9990; $s_{cv\text{-}loo}$=0.06; $F_{cv\text{-}loo}$ (p)=1173 ($2.08\cdot10^{-7}$)

Ordnance compounds - Survival and reproductive success of *Green Macroalgae*

The NOEC in the survival and reproductive success of the *Green Macroalgae* of the investigated ordnance compounds proved to be of both geometrical and topological nature and related with the compounds' atomic electronegativity and partial charge (model with two descriptors). The model with two descriptors proved to have excellent estimation and prediction power.

Sample size [reference]	8 [109]	
MDF SAR Equation	$\hat{y}=-1.28\cdot x+3.71$	$\hat{y}=-0.89\cdot x_1-13455\cdot x_2+28.03$
SAR Determination (%)	92	99.9
MDF Descriptor(s): x_1 & x_2	LnDRJQt	IHMRFEg & INPmsQt
Dominant Atomic Property	Charge (Q)	Electronegativity (E) & Charge (Q)
Interaction Via	Bonds (topology)	Space (geometry) & Bonds (topology)
Interaction Model	$Q\cdot d$	$E^{2}\cdot d^{-2}$ & $Q^{2}\cdot d^{-3}$
Structure on Activity Scale	Logarithmic	Identity & Identity
Model Statistics	r=0.9578; s=0.36; F (p)=65 ($1.83\cdot10^{-4}$)	r=0.9998; s=0.03; F (p)=5938 ($3.63\cdot10^{-9}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9237; $s_{cv\text{-}loo}$=0.61; $F_{cv\text{-}loo}$ (p)=19 ($4.68\cdot10^{-3}$)	$r_{cv\text{-}loo}$=0.9994; $s_{cv\text{-}loo}$=0.05; $F_{cv\text{-}loo}$ (p)=1821 ($6.95\cdot10^{-8}$)

Ordnance compounds - *Redfish* larvae survival (1st column) & survival and reproductive success of *Polychaete* (2nd column)

The NOEC of the investigated ordnance compounds in *redfish* larvae survival proved to be of both geometrical and topological nature and related with partial charge and the number of directly bounded hydrogens.

The NOEC of the investigated ordnance compounds in the survival and reproductive success of *Polychaete* was of geometrical nature and related with partial charge.

Sample size [reference]	8 [109]		6 [109]
MDF SAR Equation	$\hat{y}=-1.37\cdot x+0.09$	$\hat{y}=-0.003\cdot x_1+2.25\cdot x_2+4.59$	$\hat{y}=-1.42\cdot x-10.25$
SAR Determination (%)	91	99.9	95
MDF Descriptor(s): x_1 & x_2	LHDmjQg	asDmkQg & IGMmTHt	LsmrfQg
Dominant Atomic Property	Charge (Q)	Charge (Q) & Hydrogen (H)	Charge (Q)
Interaction Via	Space (geometry)	Space (geometry) & Bonds (topology)	Space (geometry)
Interaction Model	$Q^{-1}\cdot d^{-1}$	$Q^{-2}\cdot d^{-1}$ & $H^{2}\cdot d^{-4}$	$Q^{2}\cdot d^{-2}$
Structure on Activity Scale	Logarithmic	Inversed & Identity	Logarithmic
Model Statistics	r=0.9542; s=0.24; F (p)=61 ($2.33\cdot10^{-4}$)	r=0.9995; s=0.03; F (p)=2373 ($3.59\cdot10^{-8}$)	r=0.9754; s=0.32; F (p)=78 ($8.98\cdot10^{-4}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9162; $s_{cv\text{-}loo}$=0.34; $F_{cv\text{-}loo}$ (p)=28 ($1.84\cdot10^{-3}$)	$r_{cv\text{-}loo}$=0.9987; $s_{cv\text{-}loo}$=0.05; $F_{cv\text{-}loo}$ (p)= 907 ($3.96\cdot10^{-7}$)	$r_{cv\text{-}loo}$=0.9519; $s_{cv\text{-}loo}$=0.46; $F_{cv\text{-}loo}$ (p) = 37 ($3.70\cdot10^{-3}$)

Ordnance compounds - Juveniles survival of *Opossum Shrimp*

The NOEC of the investigated ordnance compounds in the juvenile survival of the *Opossum Shrimp* was of geometrical nature and related with partial charge and cardinality (model with two descriptors). The determination coefficient of the model with one descriptor was moderately good while the determination power of the model with two descriptors was very good.

Sample size [reference]	8 [109]	
MDF SAR Equation	$\hat{y}=668.36\cdot x+19.24$	$\hat{y}=1.46\cdot x_1-0.008\cdot x_2+0.28$
SAR Determination (%)	82	99.9
MDF Descriptor(s): x_1 & x_2	iBPMwEt	iIPdqQg & iImrSCg
Dominant Atomic Property	Electronegativity (E)	Charge (Q) & Cardinality (C)
Interaction Via	Bonds (topology)	Space (geometry) & Space (geometry)
Interaction Model	$E^{2}\cdot d^{-1}$	$Q^{-1}(\sqrt{Q})^{-1}$ & $C^{2}\cdot d^{-3}$
Structure on Activity Scale	Inversed	Inversed & Inversed
Model Statistics	r=0.9048; s=0.28; F (p)=27 ($2.01\cdot10^{-3}$)	r=0.9998; s=0.01; F (p)=5435 ($4.53\cdot10^{-9}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8522; $s_{cv\text{-}loo}$=0.40; $F_{cv\text{-}loo}$ (p)=11 ($1.75\cdot10^{-2}$)	$r_{cv\text{-}loo}$=0.9993; $s_{cv\text{-}loo}$=0.03; $F_{cv\text{-}loo}$ (p)=1695 ($8.32\cdot10^{-8}$)

Lowest observed effect concentration (LOEC)

Ordnance compounds - fertilization (1st column) & - embryological development (2nd column) of *Sea Urchin*

The LOEC of the investigated ordnance compounds in the fertilization and embryological development of the *Sea Urchin* was of topological nature and depended on partial charge. The estimation power of the models was good, the values were higher than or equal with 93%.

Sample size [reference]	6 [109]	7 [109]
MDF SAR Equation	$\hat{y}$=-47.56·x+0.57	$\hat{y}$=-1.14·x-7.62
SAR Determination (%)	99.8	93
MDF Descriptor(s)	IAPmfQt	lNPmfQt
Dominant Atomic Property	Charge (Q)	Charge (Q)
Interaction Via	Bonds (topology)	Bonds (topology)
Interaction Model	$Q^2 \cdot d^{-2}$	$Q^2 \cdot d^{-2}$
Structure on Activity Scale	Identity	Logarithmic
Model Statistics	r=0.9993; s=0.04; F (p)=2781 ($7.74 \cdot 10^{-7}$)	r=0.9653; s=0.36; F (p)=68 ($4.22 \cdot 10^{-4}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9981; $s_{cv\text{-}loo}$=0.10; $F_{cv\text{-}loo}$ (p)=479 ($2.58 \cdot 10^{-5}$)	$r_{cv\text{-}loo}$=0.9356; $s_{cv\text{-}loo}$=0.49; $F_{cv\text{-}loo}$ (p)=35 ($1.99 \cdot 10^{-3}$)

Ordnance compounds - Germination of *Sea Urchin*

The LOEC of the investigated ordnance compounds in the germination of the *Sea Urchin* was both of topological and geometrical nature. It was also related with group electronegativity and the relative atomic mass. The estimation and prediction power of this model was much closer to the optimum value.

Sample size [reference]	8 [109]	
MDF SAR Equation	$\hat{y}$=0.06·x-1.43	$\hat{y}$=11.77·x_1+14.55·x_2-76.46
SAR Determination (%)	88	99.9
MDF Descriptor(s): x_1 & x_2	aIDmjQg	lGPrfGt & iGPMqMg
Dominant Atomic Property	Charge (Q)	Group Electronegativity (G) & Mass (M)
Interaction Via	Space (geometry)	Bonds (topology) & Space (geometry)
Interaction Model	$Q^{-1} \cdot d^{-1}$	$E^2 \cdot d^{-2}$ & $M^{-1}(\sqrt{M})^{-1}$
Structure on Activity Scale	Inversed	Logarithmic & Inversed
Model Statistics	r=0.9357; s=0.40; F (p)=42 ($6.33 \cdot 10^{-4}$)	r=0.9996; s=0.04; F (p)=3002 ($1.99 \cdot 10^{-8}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9022; $s_{cv\text{-}loo}$=0.51; $F_{cv\text{-}loo}$ (p)=24 ($2.73 \cdot 10^{-3}$)	$r_{cv\text{-}loo}$=0.9992; $s_{cv\text{-}loo}$=0.05; $F_{cv\text{-}loo}$ (p)=1492 ($1.14 \cdot 10^{-7}$)

Ordnance compounds - Germling length and cell number of *Green Macroalgae*

The LOEC of the investigated ordnance compounds in the germling length and cell number of the *Green Macroalgae* revealed to be of topological nature and strongly depended on atomic electronegativity (model with two descriptors).

Sample size [reference]	8 [109]	
MDF SAR Equation	$\hat{y}=0.06\cdot x-2.02$	$\hat{y}=0.66\cdot x_1-688.62\cdot x_2-0.47$
SAR Determination (%)	90	99.9
MDF Descriptor(s): x_1 & x_2	aIDmjQg	ISPRfEt & imDrwEt
Dominant Atomic Property	Charge (Q)	Electronegativity (E) & Electronegativity (E)
Interaction Via	Space (geometry)	Bonds (topology) & Bonds (topology)
Interaction Model	$Q^{-1}\cdot d^{-1}$	$E^2\cdot d^{-2}$ & $E^2\cdot d^{-1}$
Structure on Activity Scale	Inversed	Identity & Inversed
Model Statistics	r=0.9504; s=0.35; F (p) = 56 ($2.94\cdot10^{-4}$)	r=0.9995; s=0.04; F (p) 2539 ($3.03\cdot10^{-8}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9320; $s_{cv\text{-}loo}$=0.41; $F_{cv\text{-}loo}$ (p)=39 ($7.99\cdot10^{-4}$)	$r_{cv\text{-}loo}$=0.9990; $s_{cv\text{-}loo}$=0.05; $F_{cv\text{-}loo}$ (p)=1255 ($1.76\cdot10^{-7}$)

Ordnance compounds - Survival and reproductive success of *Polychaete*

The LOEC of the investigated ordnance compounds in the survival and reproductive success of the *Polychaete* revealed to be of both geometrical and topological nature and related with atomic and group electronegativity. Both models had good estimation and prediction power.

Sample size [reference]	8 [109]	
MDF SAR Equation	$\hat{y}=16.60\cdot x-1.69$	$\hat{y}=-20.97\cdot x_1+51.18\cdot x_2-267.22$
SAR Determination (%)	92	99.9
MDF Descriptor(s): x_1 & x_2	anDRJQt	lsPmkEg & lSmRFGt
Dominant Atomic Property	Charge (Q)	Electronegativity (E) & Group Electronegativity (G)
Interaction Via	Bonds (topology)	Space (geometry) & Bonds (topology)
Interaction Model	$Q\cdot d$	$E^{-2}\cdot d^{-1}$ & $G^2\cdot d^{-2}$
Structure on Activity Scale	Inversed	Logarithmic & Logarithmic
Model Statistics	r=0.9612; s=0.34; F (p)=73 ($1.42\cdot10^{-4}$)	r=0.9999; s=0.02; F (p)=13109 ($5.02\cdot10^{-10}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9361; $s_{cv\text{-}loo}$=0.43; $F_{cv\text{-}loo}$ (p)=42 ($6.33\cdot10^{-4}$)	$r_{cv\text{-}loo}$=0.9998; $s_{cv\text{-}loo}$=0.03; $F_{cv\text{-}loo}$ (p)=5687 ($4.05\cdot10^{-9}$)

Ordnance compounds - Survival and reproductive success of *Green Macroalgae* (1st column) & *Redfish* larvae survival (2nd column) & *Juveniles* survival of *Opossum Shrimp* (3rd column)

The LOEC of the investigated ordnance compounds in the survival and reproductive success of the *Green Macroalgae* revealed to be of geometrical nature and depended on partial charge.

The LOEC of the investigated ordnance compounds in the *Redfish* larvae survival proved to be of geometrical nature and depended on partial charge.

The LOEC of the investigated ordnance compounds in the juvenile survival of the *Opossum Shrimp* revealed to be of geometrical nature. It also depended on cardinality.

All three models had great abilities in estimation and prediction, their power being higher than 88%.

Sample size [reference]	7 [109]	7 [109]	7 [109]
MDF SAR/ Equation	$\hat{y}=0.11\cdot x-3.69$	$\hat{y}=-1.30\cdot x+0.39$	$\hat{y}=-0.83\cdot x+4.22$
SAR Determination (%)	95	94	98
MDF Descriptor(s)	iIDdPQg	LHDmjQg	imMrtCg
Dominant Atomic Property	Charge (Q)	Charge (Q)	Cardinality (C)
Interaction Via	Space (geometry)	Space (geometry)	Space (geometry)
Interaction Model	Q^2	$Q^{-1}\cdot d^{-1}$	$C^2\cdot d^{-4}$
Structure on Activity Scale	Inversed	Logarithmic	Inversed
Model Statistics	r=0.9764; s=0.28; F (p)=102 ($1.62\cdot10^{-4}$)	r=0.9694; s=0.20; F (p)=78 ($3.09\cdot10^{-4}$)	r=0.9897; s=0.07; F (p)=239 ($2.06\cdot10^{-5}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9534; $s_{cv\text{-}loo}$=0.41; $F_{cv\text{-}loo}$ (p)=45 ($1.09\cdot10^{-3}$)	$r_{cv\text{-}loo}$=0.9404; $s_{cv\text{-}loo}$=0.29; $F_{cv\text{-}loo}$ (p)=34 ($2.08\cdot10^{-3}$)	$r_{cv\text{-}loo}$=0.9790; $s_{cv\text{-}loo}$=0.10; $F_{cv\text{-}loo}$ (p)=111 ($1.33\cdot10^{-4}$)

Discussions

The researchers from many chemistry related fields are interested in quantitative structure-activity/property relationship approaches due to their advantages. A quick search in the SCOPUS database (http://www.scopus.com) with the query “TITLE-ABS-KEY(qsar)”, retrieved a number of 6278 abstracts, out of which almost 10% were published last year. The same query retrieved a number of 4920 items in the PubMed database (http://www.ncbi.nlm.nih.gov). Out of these, almost 13% where published last year and almost 26% in the last two years. The amount of

research published on the QSAR subject showed the importance of these studies. The main three advantages offered by the approaches are as follows:

÷ Time and money efficiency [111]
÷ Effective (with comparable or better accuracy) alternatives compared with experimental counterparts [111]
÷ Virtual screening environments (particularly receptor-based virtual screening) offered reliable and inexpensive methods for identifying leads [112].

A new approach termed MDF SPR/SAR that used the information extracted from the 2D and 3D structure of the compounds in order to generate and calculate the molecular descriptors able to estimate and predict property/activity of interest was developed.

The approach was tested on different properties and activities in a number of chemical and/or biological active classes of compounds.

The analysis of the MDF abilities in estimation and prediction of properties lead to the following:

÷ The estimated power expressed as correlation coefficient varied from 0.8324 to 1.0000.
÷ The prediction power expressed as leave-one-out correlation coefficient varied from 0.8258 to 0.9999.
÷ The stability of the model, expressed as the difference between estimated and predictive power, varied from 0.01% to 5.46%. The most stable models (with a difference between estimated and predicted power of 0.0001) were obtained in the following classes of compounds: polychlorinated biphenyls (relative retention time), organophosphorus herbicides (retention chromatography index), cyclic organophosphorus (molar refraction), alkanes (boiling point), and standard amino acids – 15aa (Hückel energy).
÷ The estimation power of the best performing model (when two models were reported) occurred in 90% of cases (eighteen out of twenty investigated sets) higher than or equal with 0.9.
÷ The prediction power of the best performing model (when two models were reported) occurred in 85% of cases (seventeen out of twenty investigated sets) higher than or equal with 0.9.

The analysis of the MDF abilities in estimation and prediction of activities leads to the following:

- The estimated power expressed as correlation coefficient varied from 0.6649 to 0.9999.
- The predictive power expressed as leave-one-out correlation coefficient varied from 0.5961 to 0.9999.
- The stability of the model, expressed as the difference between estimation and prediction power, varied from 0.01% to 7.92%. The most stable model (with a difference between estimated and predicted power of 0.0001) was obtained when the fertilization of the *Sea Urchin* was investigated. The sample size of this set was small (8 compounds). A lower difference between estimation and prediction power (of 0.0002) was also obtained in the investigation of antioxidant efficacy of 3-indolyl derivates. The worst result was obtained in the toxicity investigation of the mono-substituted nitrobenzenes, where a difference of 0.0792 was obtained.
- The estimated power of the best performing model (when two models were reported) occurred in almost 77% of cases (fifty-six out of seventy-three investigated sets) higher than or equal with 0.9.
- The predicted power of the best performing model (when two models were reported) occurred in 70% of cases (fifty-one out of seventy-three investigated sets) higher than or equal with 0.9.

The MDF SPR/SAR models proved to have good correlation coefficient and predictive power compared with the previously reported models [23,29,32,34,35,47,54,56,59,67,88,90,92,94,96,99,101,103,105,106,107]. The MDF methodology was successful in extracting information from the 2D and 3D structure of compounds. This is useful for identifying the link between the compounds' structure and property or activity of interest. The use of structural information to characterize chemical compounds allows good correlations between the compounds' structure and chemical and/or biological activity. Thus, the MDF methodology is a powerful approach that investigates the activities and properties of chemical active compounds since it includes the information extracted from the compounds' structure in the SPR/SAR models. Furthermore, the approach could be useful in the process of drug design.

References

1. Hammett, L.P. 1937, J. Am. Chem. Soc., 59, 96.
2. Hansch, C., Leo, A., and Taft, R.W. 1991, Chem. Rev., 91, 165.
3. Crum-Brown, A., and Fraser, T.R. 1868, Philos. Trans. R. Soc. Lond., 25, 151.
4. Richet, C.R. 1893, C. R. Seances Soc. Biol. Fil., 9, 775.

5. Meyer, H. 1899, Naunyn-Schmiedeberg's Arch. Pharmacol., 42, 109.
6. Hammett, L.P. 1935, Chem. Rev., 17, 125.
7. Taft, R.W. 1952, J. Am. Chem. Soc., 74, 3120.
8. Hansch, C. 1969, Acc. Chem. Res., 2, 232.
9. Mezey, P.G., Journal of Mathematical Chemistry, Springer Netherlands, since 1987 (periodical).
10. Gutman, I., MATCH - Communications in Mathematical and in Computer Chemistry, Faculty of Science Kragujevac and University of Kragujevac Serbia, from 1975 (periodical).
11. Brändas, E., and Öhrn, Y., International Journal of Quantum Chemistry, John Wiley & Sons NY USA, since 1967 (periodical).
12. Portoghese, P.S., Journal of Medicinal Chemistry, American Chemical Society OH USA, since 1959 (periodical).
13. Atta-ur-Rahman. Letters in Drug Design & Discovery, Bentham Science Publishers, since 2004 (periodical).
14. Sawyer, T.K., Chemical Biology & Drug Design, Blackwell Publishing, since 2006 (periodical, formally known as Journal of Peptide Research until 2005).
15. Grassy, G., Calas, B., Yasri, A., Lahana, R., Woo, J., Iyer, S., Kaczorek, M., Floc'h, R., and Buelow, R. 1998, Nat. Biotechnol., 16, 748.
16. van Breemen, R.B. Combinatorial Chemistry & High Throughput Screening, Bentham Science Publishers, from 1998 (periodical).
17. Czarnik AW. Journal of Combinatorial Chemistry, American Chemical Society OH USA, from 1999 (periodical).
18. Hansch, C., and Leo, A. 1979, Substituent Constants for Correlation Analysis in Chemistry and Biology, John Wiley & Sons, Inc., New York.
19. Cramer, R.D. III, Patterson, D.E., and Bunce, J.D. 1988, J. Am. Chem. Soc., 110, 5959.
20. Todeschini, R., Lasagni, M., and Marengo, E. 1994, Theory J. Chemometrics, 8, 263.
21. Simon, Z., Chiriac, A., Holban, S., Ciubotariu, D., and Mihalas, G.I. 1984, Minimum Steric Difference. The MTD Method for QSAR Studies, Research Studies Press, 1.
22. Jäntschi, L., Katona, G., and Diudea, M.V. 2000, MATCH Commun. Math. Comput. Chem., 41, 151.
23. Jäntschi, L. 2004, Leonardo Journal of Sciences, 3, 68.
24. Putz, M.V., and Lacrămă, A.-M. 2007, Int. J. Mol. Sci., 8, 363.
25. Cambridge Structural Database (CSD), available from: http://www.ccdc.cam.ac.uk/products/csd/; Protein Data Bank (PDB), available from: http://www.pdb.org/; Visual Molecular Dynamics (VMD), available from: http://www.ks.uiuc.edu/Research/vmd/; Molecular Modeling DataBase (MMDB), available from: http://130.14.29.110/Structure/MMDB/mmdb.shtml; PubChem Compounds (PCC), available from: http://ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pccompound; PubChem Substance (PCS). available from: http://ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pcsubstance.
26. Eriksson, L., Jaworska, J., Worth, A.P., Cronin, M.T.D., McDowell, R.M., and Gramatica, P. 2003, Environ. Health Perspect., 111, 1361.

27. Steiger, J.H. 1980, Psychol. Bull., 87, 245.
28. Thurstone, L.L. 1931, Psychol. Rev., 38, 406.
29. Bolboacă, S.D., and Jäntschi, L. 2006, Leonardo Journal of Sciences, 5, 179.
30. Tarko, L. 2006, Revista de Chimie, 57, 1014.
31. Jäntschi, L., Bolboacă, S.D. 2007, Leonardo Electronic Journal of Practices and Technologies, 6, 169.
32. Jäntschi, L. 2005, Leonardo Electronic Journal of Practices and Technologies, 6, 76.
33. HyperChem, Molecular Modelling System [online]; ©2003, Hypercube [cited 2007 Feb]. Available from: URL: http://hyper.com/products/.
34. Bolboacă, S.D., and Jäntschi, L. 2008, Environ. Chem. Lett., 6, 175.
35. Jäntschi, L., and Bolboacă, S.D. 2007, Studii şi Cercetări Ştiinţifice Universitatea Bacău Seria Biologie, 12, 57.
36. Lu. A., Zhang. J., Yin. X., Luo. X., and Jiang. H. 2007, Bioorg. Med. Chem. Lett., 17, 243.
37. Ma, W., Luan, F., Zhao, C., Zhang, X., Liu, M., Hu, Z., and Fan, B., 2006, QSAR Comb. Sci., 25, 895.
38. Tropsha, A., Gramatica, P., and Gombar, V.K. 2003, QSAR & Comb. Sci., 22, 69.
39. Diudea, M., Gutman, I., and Jäntschi, L. 2002, Molecular Topology, 2nd Edition, Nova Science, Huntington, New York.
40. The PHP Group [online]; ©2001-2007, The PHP Group [cited 2007 June]. Available from: URL: http://php.net.
41. MySQL AB [online]; ©1995-2007 MySQL AB [cited 2007 June]. Available from: URL: http://mysql.com.
42. The FreeBSD Project [online]; ©1995-2007 The FreeBSD Project [cited 2007 June]. Available from: URL: http://freebsd.org.
43. Chambers, D.L. 2001, The practical handbook of genetic algorithms. Chapman & Hall, Boca Raton.
44. Borland Software Corporation [online]; ©1994 - 2007 Borland Software Corporation [cited 2007 June]. Available from: URL: http://borland.com.
45. Hawkins, D.M. 2004, J. Chem. Inf. Comput. Sci., 44, 1.
46. Abraham, M. H., Kumarsingh, R., Cometto-Muniz, J.E., and Cain, W. S. 1998, Toxicol. In Vitro., 12, 201.
47. Ivanciuc, O. 1998, Revue Roumanian de Chimie, 43, 255.
48. Jäntschi, L., and Bolboacă, S.D. 2007, Int. J. Quantum Chem., 107, 1736.
49. Eisler, R., and Belisle, A.A. 1996, Biological Report 31 and Contaminant Hazard Reviews Report, 31, 75.
50. Baker, J.R., Mihelcic, J.R., and Sabljie, A. 2001, Chemosphere, 45, 213.
51. Bumble, S. 1999, Computer Generated Physical Properties. CRC Press, LLC, Boca Raton.
52. Bolboacă, S.D., and Jäntschi, L. 2008, Chem. Biol. Drug Des., 71(2), 173.
53. Jäntschi, L., Mureşan, S., and Diudea, M. 2000, Studia Universitatis Babes-Bolyai, Chemia, XLV, 313.
54. Jäntschi, L., and Bolboacă, S.D. 2005, Leonardo Electronic Journal of Practices and Technologies, 7, 55.
55. Toropov, A., Toropova, A., Ismailov, T., and Bonchev, D. 1998, Theochem, 424, 237.

56. Brasquet, C., and Le Cloirec, P. 1999, Water Res., 33, 3603.
57. Jäntschi, L. 2004, Leonardo Journal of Sciences, 5, 63.
58. Wolfenden, R., Andersson, L., Cullis, P., and Southgate, C. 1981, Biochemistry, 20, 849.
59. Rose,G.D., Geselowitz, A.R., Lesser, G.J., Lee, R.H., and Zehfus, M.H., 1985, Science, 229, 834.
60. Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S.H., and von Heijne, G. 2005, Nature, 433, 377.
61. Kyte, J., and Doolittle, 1982, R.F., J. Mol. Biol., 157, 105.
62. Welling, G.W., Weijer, W.J., Van der Zee, R., and Welling-Wester, S. 1985, FEBS Lett, 188, 215.
63. Wilson, K.J., Honegger, A., Stotzel, R.P., and Hughes, G.J. 1981, Biochem. J., 199, 31.
64. Cornette, J.L., Cease, K.B., Margalit, H., Spouge, J.L., Berzofsky, J.A., and DeLisi, C. 1987, J. Mol. Biol., 195, 659.
65. Bolboacă, S.D., and Jäntschi, L. 2007, Recent Advances in Synthesis & Chemical Biology VI, Centre for Synthesis & Chemical Biology, University of Dublin, December 14, Dublin, Ireland, P2.
66. Wimley, W.C., and White, S.H. 1996, Nature Struct. Biol., 3, 842.
67. Hoop, T.P., and Woods, K.R. 1981, Proc. Natl. Acad. Sci. USA, 78, 3824.
68. Cowan, R. 1990, Pept. Res., 3, 75.
69. Manavalan, P., and Ponnuswamy, P.K. 1978, Nature, 275, 673.
70. Fauchere, J.-L., and Pliska, V.E. 1983, Eur. J. Med. Chem., 18, 369.
71. Rao, M.J.K., and Argos, P. 1986, Biochim. Biophys. Acta., 869, 197.
72. Janin J, Surface and inside volumes in globular proteins, Nature, 1979, 277, 491-492.
73. Roseman., M.A. 1988, J. Mol. Biol., 200, 513.
74. Urry, D.W., 2004, Chem. Phys. Lett., 399, 177.
75. Engelman, D.M., Steitz, T.A., and Goldman, A., 1986, Annu. Rev. Biophys. Chem., 15, 321.
76. Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. 1984, J. Mol. Biol., 179, 125.
77. Sereda, T.J., Mant, C.T., Sonnichsen, F.D., and Hodges, R.S. 1994, J. Chromatogr. A., 676, 139.
78. Bull, H.B., and Breese, K. 1974, Arch. Biochem. Biophys., 161, 665.
79. Parker, J.M.R, Guo, D., and Hodges, R.S., 1986, Biochemistry, 25, 5425.
80. Black, S.D., Mould, D.R., Black, S.D., and Mould, D.R. 1991, Anal. Biochem., 193, 72.
81. Monera, O.D., Sereda, T.J., Zhou, N.E., Kay, C.M., and Hodges, R.S. 1995, J. Pept. Sci., 1, 319.
82. Wei, D., Zhang, A., Wu, C., Han, S., and Wang, L. 2001, Chemosphere, 44, 1421.
83. Agrawal, V.K., and Khadikar, P.V. 2001, Bioorg. Med. Chem., 9, 3035.
84. Toropov, A.A., and Toropova, A.P. 2002, Theochem, 578, 129.
85. Ade, T., Zaucke, F., and Krug, H.F. 1996, Fresenius J. Anal. Chem., 354, 609.
86. Bobloacă, S.D., and Jäntschi, L. 2006, Terapeutics, Pharmacology and Clinical Toxicology, X(1), 110.
87. Smith, C.J., Hansch, C., and Morton, M.J. 1997, Mutat. Res., 379, 167.

88. Jäntschi, L., and Bolboacă, S. 2005, The 10th Electronic Computational Chemistry Conference, April 2005.
89. Hasegawa, K., Arakawa, M., and Funatsu, K. 1999, Chemometr. Intell. Lab., 47, 33.
90. Bolboacă, S., and Jäntschi, L. 2005, Leonardo Journal of Sciences, 6, 78.
91. Diudea, M., Jäntschi, L., and Pejov, L. 2002, Leonardo Electronic Journal of Practices and Technologies, 1, 1.
92. Bolboacă, S., and Jäntschi, L. 2006, Bulletin of University of Agricultural Sciences and Veterinary Medicine – Agriculture, 62, 35.
93. Shertzer,G.H., Tabor, M.W., Hogan, I.T.D., Brown, J.S., and Sainsbury, M. 1996, Arch. Toxicol., 70, 830.
94. Bolboacă, S., Filip, C., Ţigan, Ş., and Jäntschi, L. 2006, Clujul Medical, LXXIX, 204.
95. Zhou, Y.X., Xu, L., Wu, Y.P., and Liu, B.L. 1999, Chemometr. Intell. Lab., 45, 95.
96. Frahm, A.W., and Chaudhuri, R. 1979, Tetrahedon, 35, 2035.
97. Bolboacă, S.D., and Jäntschi, L. 2005, Leonardo Journal of Sciences, 7, 58-64.
98. Morita, H., Gonda, A., Wei, L., Takeya, K., and Itokawa, H. 1997, Bioorg. Med. Chem. Lett., 7, 2387-2392.
99. Bolboacă, S.D., and Jäntschi, L. 2008, Arch. Med. Sci., 4(1), 7.
100. Castro, E.A., Torrens, F., Toropov, A.A., Nesterov, I.V., and Nabiev, O.M. 2004, Mol. Simulat., 30, 691.
101. Bolboacă, S., Ţigan, Ş., and Jäntschi, L. 2006, Proceedings of the European Federation for Medical Informatics Special Topic Conference, April 6-8, 222.
102. Supuran, C.T., and Clare, B.W. 1999, Eur. J. Med. Chem., 34, 41.
103. Bolboacă, S.D., and Jäntschi, L. 2007, Integration of Structure Information, Computer-Aided Chemical Engineering, Elsevier Netherlands & UK, 24, 965.
104. Jäntschi, L., Ungureşan, M.L., and Bolboacă, S.D. 2005, Applied Medical Informatics, 17, 12.
105. Jäntschi, L., Bolboacă, S. 2006, Electron. J. Biomed., 2, 22.
106. Opris, D., and Diudea, M.V. 2001, SAR QSAR Environ. Res., 12, 1599.
107. Selassie, C.D., Li, R.L., Poe, M., and Hansch, C. 1991, J. Med. Chem., 34, 46.
108. Hellberg, S., Eriksson, L., Jonsson, J., Lindgren, F., Sjostrom, M., Skagerberg, B., Wold, S., and Andrews, P. 1991, Int. J. Pept. Protein Res., 37, 414.
109. Carr, R.S., and Nipper, M. 2000. Toxicity of marine sediments and porewaters spiked with ordnance compounds: Report prepared for Naval Facilities Engineering Commands, NFESC Contract Report CR 01-001-ENV, Washington, DC, USA. Available from: http://enviro.nfesc.navy.mil/erb/erb_a/restoration/fcs_area/con_sed/MarineSed2000.pdf.
110. Stoenoiu, C.E., Bolboacă, S.D., and Jäntschi, L. 2007, MetEcoMat - Ecomaterials and Processes: Characterization and Metrology, April 19-21, St. Kirik, Plovdiv, Bulgaria, 54.
111. Prathipati, P., Dixit, A., and Saxena, A.K. 2007, Current Computer-Aided Drug Design, 3, 133.
112. Lyne, P.D. 2002, Drug Discov. Today, 7, 1047.