"Iuliu Haţeiganu" University of Medicine and Pharmacy Cluj-Napoca

Quantitative Structure-Activity Relationships as Prediction Tools

qSAR

Contents

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, 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], Quantum Chemistry [10], and Medicinal Chemistry [11] 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 [12].

The aim of the present paper is to present some models with abilities in prediction of a series of chromatographic parameters and of hydrophobic/hydrophilic character.

Structure- Property/Activity Relationships: Models by Examples SPR Models

In order to justify the introduction of the molecular descriptors family on the structure-property relationships the chromatogtaphic parameters on classes of compounds have been investigated and are presented.

Chromatographic Parameters
Polychlorinated Biphenyls - Relative Retention Time

Sample size	206	
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 & IADrtHg
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 ($1.64 \cdot 10^{-189}$)	r=0.9986; s=0.01; F (p)=36600 ($1.10 \cdot 10^{-265}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9920; $s_{cv\text{-}loo}$=0.02; $F_{cv\text{-}loo}$ (p)=12777 ($1.06 \cdot 10^{-188}$)	$r_{cv\text{-}loo}$=0.9985; $s_{cv\text{-}loo}$=0.01; $F_{cv\text{-}loo}$ (p)=35416 ($3.33 \cdot 10^{-264}$)

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).

Polychlorinated Biphenyls - Relative Response Factor

Sample size	209	
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)
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-loo}=0.7873; s_{cv-loo}=0.22; F_{cv-loo} (p)=337 ($2.02 \cdot 10^{-45}$)	r_{cv-loo}=0.8258; s_{cv-loo}=0.20; F_{cv-loo} (p)=221 ($3.73 \cdot 10^{-52}$)

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.

Organophosphorus Herbicides - Retention Chromatography Index

Sample size	10	
MDF SPR Equation	$\hat{y}=0.32 \cdot x-3.37$	$\hat{y}=6.37 \cdot x_1+0.06 \cdot x_2-62.36$
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/HvH	$M^2 \cdot d^{-1}$ & $M \cdot d^{-2}$
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-loo}=0.9563; s_{cv-loo}=0.95; F_{cv-loo} (p)=85 ($1.55 \cdot 10^{-5}$)	r_{cv-loo}=0.9993; s_{cv-loo}=0.13; F_{cv-loo} (p)=2344 ($1.28 \cdot 10^{-10}$)

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.

SAR Models

The abilities of the molecular descriptors family approach have been investigated on samples of biological active compounds in order to investigate their hydrophobicity/hydrophilicity.

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	15	15	15
MDF SAR Equation	$\hat{y}=-160 \cdot x - 0.07$	$\hat{y}=8.5 \cdot x - 0.58$	$\hat{y}=-21 \cdot x + 12$
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	20	20	20
MDF SAR Equation	ŷ=0.39·x-1.23	ŷ=-27.79·x+6.55	ŷ=-1.73·x-2.88
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

Sample size	20	20	20
MDF SAR Equation	ŷ=7.35·x-3.37	ŷ=10.63·x-1.99	ŷ=-6.57·x+1.47
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 ($3.00 \cdot 10^{-6}$)	r=0.8608; s=1.01; F (p)=52 ($1.11 \cdot 10^{-6}$)	r=0.8661; s=0.6; F (p)=54 ($1.15 \cdot 10^{-8}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.800; $s_{cv\text{-}loo}$=0.54; $F_{cv\text{-}loo}$ (p) = 32 ($2.25 \cdot 10^{-5}$)	$r_{cv\text{-}loo}$=0.8288; $s_{cv\text{-}loo}$=1.11; $F_{cv\text{-}loo}$ (p) = 39 ($6.49 \cdot 10^{-6}$)	$r_{cv\text{-}loo}$=0.8344; $s_{cv\text{-}loo}$=0.73; $F_{cv\text{-}loo}$ (p) = 41 ($7.94 \cdot 10^{-8}$)

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%.

Twenty Standard Amino Acids – Hydrophobicity

Sample size	20	20	20
MDF SAR Equation	ŷ=23.43·x+14.55	ŷ=5.94·x-4.36	ŷ=-2.73·x+1.43
SAR Determination (%)	78	78	79
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 ($2.84 \cdot 10^{-7}$)	r=0.8832; s=0.50; F (p)=65 ($2.50 \cdot 10^{-7}$)	r=0.8901; s=0.24; F (p)=69 ($1.48 \cdot 10^{-7}$)
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8546; $s_{cv\text{-}loo}$=0.84; $F_{cv\text{-}loo}$ (p)=49 ($1.65 \cdot 10^{-6}$)	$r_{cv\text{-}loo}$=0.8611; $s_{cv\text{-}loo}$=0.54; $F_{cv\text{-}loo}$ (p) = 51 ($1.13 \cdot 10^{-6}$)	$r_{cv\text{-}loo}$=0.8545; $s_{cv\text{-}loo}$=0.28; $F_{cv\text{-}loo}$ (p) = 48 ($1.78 \cdot 10^{-6}$)

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

Sample size	20	20	20
MDF SAR Equation	ŷ=1.74·x+0.86	ŷ=-3.78·x+5.30	ŷ=1.75·x+0.86
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}	$Q \cdot d$	$Q^2 \cdot d^{-3}$
Structure on Activity Scale	Inversed	Identity	Inversed
Model Statistics	r=0.8974; s=0.05; $F (p)=76 (6.76 \cdot 10^{-8})$	r=0.9208; s =0.80; $F (p)=100 (8.69 \cdot 10^{-9})$	r=0.8974; s=0.05; $F (p)=74 (8.21 \cdot 10^{-8})$
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8744; $s_{cv\text{-}loo}$=0.06; $F_{cv\text{-}loo} (p) = 56 (6.37 \cdot 10^{-7})$	$r_{cv\text{-}loo}$=0.9073; $s_{cv\text{-}loo}$=0.68; $F_{cv\text{-}loo} (p) = 84 (3.48 \cdot 10^{-8})$	$r_{cv\text{-}loo}$=0.8744; $s_{cv\text{-}loo}$=0.06; $F_{cv\text{-}loo} (p) = 58 (4.73 \cdot 10^{-7})$

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%).

Twenty Standard Amino Acids – Hydrophobicity

Sample size	20	20	20
MDF SAR Equation	ŷ=-11.96·x-29.73	ŷ=-753.09·x+ 1.85	ŷ=-0.92·x+ 1.68
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; $F (p)=81 (4.40 \cdot 10^{-8})$	r=0.9116; s=2.07; $F (p)=89 (2.26 \cdot 10^{-8})$	r=0.9128; s=0.42; $F (p)=90 (2.02 \cdot 10^{-8})$
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.8819; $s_{cv\text{-}loo}$=1.18; $F_{cv\text{-}loo} (p) = 63 (2.85 \cdot 10^{-7})$	$r_{cv\text{-}loo}$=0.8731; $s_{cv\text{-}loo}$=2.56; $F_{cv\text{-}loo} (p) = 51 (1.13 \cdot 10^{-6})$	$r_{cv\text{-}loo}$=0.8935; $s_{cv\text{-}loo}$=0.46; $F_{cv\text{-}loo} (p) = 70 (1.31 \cdot 10^{-7})$

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.

Twenty Standard Amino Acids – Hydrophobicity

Sample size	20	20	20
MDF SAR Equation	$\hat{y}=-2.16 \cdot x+4.64$	$\hat{y}=817.95 \cdot x+81.72$	$\hat{y}=7.18 \cdot x-0.41$
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)
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; $s_{cv\text{-}loo}$=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 $(6.01 \cdot 10^{-8})$

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. 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

Sample size	20	20	20
MDF SAR Equation	$\hat{y}=-0.20 \cdot x+1.36$	$\hat{y}=1.85 \cdot x+11.05$	$\hat{y}=19.17 \cdot x-7.60$
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	$Q \cdot d$	Q	$d \cdot Q$
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 $(7.25 \cdot 10^{-9})$

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.

Twenty Standard Amino Acids – Hydrophobicity

Sample size	20	19 (- Proline)
MDF SAR Equation	$\hat{y}=-0.96 \cdot x+0.86$	$\hat{y}=843.88 \cdot x+86.05$
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 $(1.09 \cdot 10^{-9})$	r=0.9504; s=16.49; F (p)=159 $(4.77 \cdot 10^{-10})$
Cross-Validation Leave-One-Out	$r_{cv\text{-}loo}$=0.9263; $s_{cv\text{-}loo}$=0.13; $F_{cv\text{-}loo}$ (p)=109 $(4.73 \cdot 10^{-9})$	$r_{cv\text{-}loo}$=0.9380; $s_{cv\text{-}loo}$=18.37; $F_{cv\text{-}loo}$ (p)=125 $(3.00 \cdot 10^{-9})$

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.

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] Brändas, E., and Öhrn, Y., International Journal of Quantum Chemistry, John Wiley & Sons NY USA, since 1967 (periodical).

[11] Portoghese, P.S., Journal of Medicinal Chemistry, American Chemical Society OH USA, since 1959 (periodical).

[12]Atta-ur-Rahman. Letters in Drug Design & Discovery, Bentham Science Publishers, since 2004 (periodical).