MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **60** (2008) 1021-1032

ISSN 0340 - 6253

MODELLING ANALYSIS OF AMINO ACIDS HYDROPHOBICITY

Sorana D. BOLBOACĂ[1] and Lorentz JÄNTSCHI[2]

[1]"Iuliu Haţieganu" University of Medicine and Pharmacy Cluj-Napoca, 400349 Louis Pasteur, Cluj-Napoca, Romania. E-mail: sbolboaca@umfcluj.ro

[2] Technical University of Cluj-Napoca, 103-105 Muncii Bvd., 400641 Cluj-Napoca, Romania. E-mail: lori@academicdirect.org

(Received April 27, 2008)

Abstract. The aim of the study was to perform a structural modelling analysis on amino acids hydrophobicity in order to identify, characterize and quantify the relationship between the structure and the considered property. A set of twenty essential amino acids (alanine, arginine, asparagine, aspartate, cysteine, glutamine, glutamate, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, and valine) was investigated by using the Molecular Descriptors Family on the Structure-Activity/Property Relationship approach. The property of interest was the hydrophobic or hydrophilic character measured on twenty-four different scales. The information extracted from the amino acids structure was used to calculate the Molecular Descriptors Family. The best performing monovariate model in terms of goodness-of-fit for each hydrophobicity scale were collected and analyzed. The resulted models were used to predict the hydrophobicity of a set of eleven non-standard amino acids (seleno-L-cysteine, pyrrolysine, lanthionine, 2-aminoisobutyric acid, dehydroalanine, gamma-aminobutyric acid, ornithine, citrulline, homocysteine, hydroxyproline, and dopamine). All identified models were statistically significant ($p < 0.0001$). An internal validation approach was applied for analyzing the validity of the obtained models. The correlation coefficient calculated between the measured and estimated hydrophobicity varied from 0.6649 (hydrophobicity reported by Welling *et al.* 1985) to 0.9504 (hydrophobicity reported by Monera *et al.*, 1995). The obtained results showed that the amino acids hydrophobicity is a property linearly related to the compounds structure. The amino acids hydrophobicity is strongly related to atomic charges through geometric interactions.

Introduction

Amino acids, the building blocks of proteins, molecules that contains amine and carboxyl functional groups, play an important roles in biology such as: synthesis of proteins [1,2], intermediates of metabolic pathways [3], neurotransmitters [4,5], antibiotics [6,7]. The standard amino acids were mostly investigated. The biochemical, bioinformatics and evolutionary studies of standard amino acids lead also to the development of online resources (e.g. Amino Acid Explorer[a]).

[a] Amino Acid Explorer. National Center for Biotechnology Information. URL: http://www.ncbi.nlm.nih.gov/Class/Structure/aa/aa_explorer.cgi

The quantitative investigation of the structure-activity relationships of amino acids is important in biological research [8] even only applied to the essential amino acids, i.e., amino acids found in biological systems [9], to the amino acids synthesized abiotically [10,11], or also to those engineered by scientists [12]. The quantitative structure-property relationship (QSPRs) methodology is a mathematical approach that links chemical structures and their activity or physico-chemical property, in a quantitative manner [13]. Development of such methods was possible due to computers and information technology progress, offering a less costly and less time consuming determination of activities or properties of chemicals [14,15]. Several properties of amino acids have been characterized by using these approaches [16-18]. Hydrophobic or hydrophilic character of an amino acid, very important in protein structure and protein-protein interactions, is one of the most studied properties. To date, many hydrophobicity scales were reported [19-24]. The differences between scales are significant: Janin (1979) and Kyte and Doolittle (1982) classify cistein as the most hydrophobic while Wolfenden *et al.* [25] while Rose *et al.* [24] do not. These differences could be explained by the fundamentals of the methods used in the construction of the scale.

The aim of this study was to identify and quantify the interrelation between different hydrophobicity scales and the structural information of standard amino acids.

Material and Method

Amino Acids Hydrophobicity

The amino acids under study were split into two different sets, one for estimation and the other for prediction. The first set, used for generating of the models, consists of twenty standard amino acids: 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). The second set comprises non-standard amino acids (formed by modifications of standard amino acids) and was used for prediction: selenocysteine (Sec), pyrrolysine (Pyl), lanthionine (Lth), 2-aminoisobutyric acid (Aib), dehydroalanine (Dhd), gamma-aminobutyric acid (Gab), ornithine (Oth), citrulline (Ciu), homocysteine (Hcy), hydroxyproline (Hyp), and dopamine (Dop).

The 3D structures of standard amino acids were used to derive the best model able to characterize their hydrophobic or hydrophilic character.

The hydrophobicity measured on twenty-four scales was the property of interest. The values of the measured hydrophobicity were taken from the literature. Table I lists the abbreviation of the set (Abb.) and the reference were the values were published (Ref.).

Table I. Amino acids hydrophobicity scales

Abb.	Ref.	Set abb.	Ref.
Hyd_01	[19]	Hyd_13	[26]
Hyd_02	[20]	Hyd_14	[27]
Hyd_03	[21]	Hyd_15	[28]
Hyd_04	[22]	Hyd_16	[29]
Hyd_05	[30]	Hyd_17	[31]
Hyd_06	[32]	Hyd_18	[33]
Hyd_07	[34]	Hyd_19	[35]
Hyd_08	[36]	Hyd_20	[37]
Hyd_09	[24]	Hyd_21	[38]
Hyd_10	[23]	Hyd_22	[39]
Hyd_11	[40]	Hyd_23	[41]
Hyd_12	[42]	Hyd_24	[43]

Molecular Descriptor Family on Structure-Property Relationships

A QSPR method, called Molecular Descriptors Family on Structure-Property Relationships (MDF-SPR) was introduced in [44] for which good estimated and predictive abilities in various classes of biologically active compounds were proved [45-47]. The method used the information extracted from the 2D and 3D structures of compounds in order to identify and quantify the link between compound's structure and property. For each compound, a series of molecular descriptors are calculated [44]. The name of each descriptor is a string seven letters (Table II, [48]) that shows the properties encoded by its construction.

Table II. Characters used by the name of the molecular descriptor

Letter	Characters	No. of all possible characters
First	I-i-A-a-L-l	6
Second	m-M-n-N-S-P-s-A-a-B-b-G-g-F-f-H-h-I-i	19
Third	m-M-D-P	4
Fourth	R-r-M-m-D-d	6
Fifth	D-d-O-o-P-p-Q-q-J-j-K-k-L-l-V-E-W-w-F-f-S-s-T-t	24
Sixth	C-H-M-E-G-Q	6
Seventh	g-t	2

For the set abbreviated as Hyd_24, the model was obtained based on nineteen amino acids, due to the absence of the proline hydrophobicity on the reference [43].

The models obtained by using the structure of the standard amino acids were used in estimation of the hydrophobic or hydrophilic character of the non-standard amino acids.

Results

The relation between hydrophobicity of essential amino acids and their structure was investigated. The monovariate model with the best goodness-of-fit was identified for each hydrophobicity scale on the standard amino acids set. The main characteristics of the models are presented in Table III.

A similarity analysis of the molecular descriptors used in modelling the standard amino acids was performed. The obtained frequency of the characters on the name of descriptors is presented in Table IV.

Table IV. Distribution of the characters in descriptors name

1st letter		2nd letter		3rd letter		4th letter		5th letter		6th letter		7th letter	
Cha.	f_a	Cha.	f_a	Cha.	f_a	Cha.	f_a	Cha.	f_a	Cha.	f_a	Cha.	f_a
A	4	A	3	D	7	d	4	F	1	E	1	g	19
i	16	B	5	m	11	m	2	K	3	Q	23	t	5
l	4	f	1	P	6	r	18	L	5				
		G	1					O	7				
		H	1					p	5				
		I	1					W	3				
		m	7										
		n	5										

Cha. = character; f_a = absolute frequency

Statistical characteristics for estimation and prediction (cross-validation leave-one-out analysis) of the models listed in Table III are presented in Table V.

TABLE III. MDF-SPRs models of amino acids hydrophobicity

Amino acid property	Hyd_01	Hyd_02	Hyd_03	Hyd_04
MDF SPR Equation	$\hat{Y} = 0.86 - 0.96 \cdot x$	$\hat{Y} = -7.60 + 19.17 \cdot x$	$\hat{Y} = -3.37 + 7.35 \cdot x$	$\hat{Y} = -0.41 + 7.18 \cdot x$
SPR Determination (%)	88	87	71	85
MDF Descriptor (x)	lAmrLQg	iGPdLQg	iBmrWQt	AmDROQg
Dominant Atomic Property	Charge (Q)	Charge (Q)	Charge (Q)	Charge (Q)
Interaction via	Space (geometry)	Space (geometry)	Bonds (topology)	Space (geometry)
Interaction Model	$d \cdot \sqrt{Q}$	$d \cdot \sqrt{Q}$	Q^2/d	Q
Structure on Property Scale	Proportional	Inversed	Inversed	Proportional

Amino acid property	Hyd_05	Hyd_06	Hyd_07	Hyd_08
MDF SPR Equation	$\hat{Y} = 81.72 + 817.95 \cdot x$	$\hat{Y} = -1.99 + 10.63 \cdot x$	$\hat{Y} = -2.88 - 1.73 \cdot x$	$\hat{Y} = 1.68 - 0.92 \cdot x$
SPR Determination (%)	85	74	69	83
MDF Descriptor (x)	inMrpQg	iMPRoQg	LmDROQg	IAMdKQg
Dominant Atomic Property	Charge (Q)	Charge (Q)	Charge (Q)	Charge (Q)
Interaction via	Space (geometry)	Space (geometry)	Space (geometry)	Space (geometry)
Interaction Model	Q^{-2}	Q^{-1}	Q	$Q^2 \cdot d$
Structure on Property Scale	Inversed	Inversed	Logarithmic	Logarithmic

Amino acid property	Hyd_09	Hyd_10	Hyd_11	Hyd_12

MDF SPR Equation	$\hat{Y} = 0.86 + 1.74 \cdot x$	$\hat{Y} = 0.48 + -137.72 \cdot x$	$\hat{Y} = 1.85 - 753.09 \cdot x$	$\hat{Y} = -3.36 + 3.76 \cdot x$
SPR Determination (%)	81	81	83	81
MDF Descriptor (x)	inMrpQg	IHPrFQt	INPrWQg	iBDdwQg
Dominant Atomic Property	Charge (Q)	Charge (Q)	Charge (Q)	Charge (Q)
Interaction via	Space (geometry)	Bonds (topology)	Space (geometry)	Space (geometry)
Interaction Model	Q^{-2}	Q^2/d^2	Q^2/d	Q^2/d
Structure on Property Scale	Inversed	Logarithmic	Logarithmic	Inversed

Amino acid property	Hyd_13	Hyd_14	Hyd_15	Hyd_16
MDF SPR Equation	$\hat{Y} = 1.36 - 0.20 \cdot x$	$\hat{Y} = 5.30 - 3.78 \cdot x$	$\hat{Y} = -1.23 + 0.39 \cdot x$	$\hat{Y} = 11.05 + 1.85 \cdot x$
SPR Determination (%)	85	85	44	86
MDF Descriptor (x)	ilPmLQt	IAmrLQg	amMRLQt	lfPROQg
Dominant Atomic Property	Charge (Q)	Charge (Q)	Charge (Q)	Charge (Q)
Interaction via	Bonds (topology)	Space (geometry)	Bonds (topology)	Space (geometry)
Interaction Model	Q·d	Q·d	Q·d	Q
Structure on Property Scale	Inversed	Logarithmic	Inversed	Logarithmic

Amino acid property	Hyd_17	Hyd_18	Hyd_19	Hyd_20
MDF SPR Equation	$\hat{Y} = 4.64 - 2.16 \cdot x$	$\hat{Y} = 14.55 + 23.43 \cdot x$	$\hat{Y} = -4.36 + 5.94 \cdot x$	$\hat{Y} = 1.43 - 2.73 \cdot x$
SPR Determination (%)	84	78	78	79
MDF Descriptor (x)	lbmdKQg	inMrpQg	ibDRPQg	AmDROQg
Dominant Atomic Property	Charge (Q)	Charge (Q)	Charge (Q)	Charge (Q)
Interaction via	Space (geometry)	Space (geometry)	Space (geometry)	Space (geometry)
Interaction Model	$Q^2 \cdot d$	Q^{-2}	Q^2	Q
Structure on Property Scale	Logarithmic	Inversed	Inversed	Proportional

Amino acid property	Hyd_21	Hyd_22	Hyd_23	Hyd_24
MDF SPR Equation	$\hat{Y} = 6.55 - 27.79 \cdot x$	$\hat{Y} = 1.47 + -6.57 \cdot x$	$\hat{Y} = -29.73 + -11.96 \cdot x$	$\hat{Y} = 86.05 + 843.88 \cdot x$
SPR Determination (%)	66	75	82	90
MDF Descriptor (x)	immRoQg	AmDROQg	iBDMkEt	inMrpQg
Dominant Atomic Property	Charge (Q)	Charge (Q)	Electronegativity (E)	Charge (Q)
Interaction via	Space (geometry)	Space (geometry)	Bonds (topology)	Space (geometry)
Interaction Model	Q^{-1}	Q	$Q^{-2} \cdot d^{-1}$	Q^{-2}
Structure on Property Scale	Inversed	Absolute	Inversed	Inversed

Q = change; d = distance

The models presented in Table III were used in order to predict the hydrophobicity of the non-standard amino acids. The predicted activity according to the hydrophobicity scale is given in Table VI. The hydrophobicity of proline on the Monera *et al*. scale was taken as 96.57.

TABLE V. MDF-SPRs models: statistical characteristics

Abb.	Regression model						Leave-one-out		
	n	r	F (p)	s	$[_{95\%}CI]_{Intercept}$ ($p_{t\text{-}Stat}$)	$[_{95\%}CI]_{Slop}$($p_{t\text{-}Stat}$)	r_{loo}	F_{loo}	s_{loo}
Hyd_01	20	0.9376	131 ($1.09 \cdot 10^{-9}$)	0.12	[0.77 – 0.94]*	[-1.14 – -0.78]*	0.9263	109 ($4.73 \cdot 10^{-9}$)*	0.13
Hyd_02	20	0.9327	120 ($2.10 \cdot 10^{-9}$)	1.11	[-9.05 – - 6.14]*	[15.50 - 22.84]*	0.9226	103 ($7.25 \cdot 10^{-9}$)	1.18
Hyd_03	20	0.8434	44 ($3.00 \cdot 10^{-6}$)	0.48	[-4.42 – - 2.32]*	[5.03 – 9.67]*	0.8009	32 ($2.25 \cdot 10^{-5}$)	0.54
Hyd_04	20	0.9238	105 ($6.24 \cdot 10^{-9}$)	0.52	[-0.79 – - 0.02]*	[5.70 – 8.65]*	0.9018	78 ($6.01 \cdot 10^{-8}$)	0.58
Hyd_05	20	0.9232	104 ($6.69 \cdot 10^{-9}$)	20.73	[66.20 – 97.23]*	[649.29-986.61]*	0.9082	85 ($3.16 \cdot 10^{-8}$)	22.58
Hyd_06	20	0.8608	52 ($1.11 \cdot 10^{-6}$)	1.01	[-2.70 – - 1.29]*	[7.52 – 13.75]*	0.8288	39 ($6.49 \cdot 10^{-6}$)	1.11
Hyd_07	20	0.8309	40 ($5.70 \cdot 10^{-6}$)	1.70	[-4.30 – - 1.39]*	[-2.30 – - 1.15]*	0.7936	30 ($3.34 \cdot 10^{-5}$)	1.87
Hyd_08	20	0.9128	90 ($2.02 \cdot 10^{-8}$)	0.42	[1.26 – 2.10]*	[-1.12 – - 0.72]*	0.8935	70 ($1.31 \cdot 10^{-7}$)	0.46
Hyd_09	20	0.8974	74 ($8.21 \cdot 10^{-8}$)	0.05	[0.82 – 0.90]*	[1.32 – 2.17]*	0.8744	58 ($4.73 \cdot 10^{-7}$)	0.06
Hyd_10	20	0.8997	76 ($6.76 \cdot 10^{-8}$)	0.32	[0.29 – 0.70]*	[-172.49 – - 105.66]*	0.8599	56 ($6.37 \cdot 10^{-7}$)	0.36
Hyd_11	20	0.9116	89 ($2.26 \cdot 10^{-8}$)	2.07	[0.64 – 3.06]*	[-921.24 – - 584.95]*	0.8731	51 ($1.13 \cdot 10^{-6}$)	2.56
Hyd_12	20	0.8986	75 ($7.42 \cdot 10^{-8}$)	0.45	[-4.22 – - 2.50]*	[2.85 – 4.67]*	0.8812	62 ($2.93 \cdot 10^{-7}$)	0.48
Hyd_13	20	0.9252	107 ($5.30 \cdot 10^{-9}$)	0.36	[1.02 – 1.70]*	[-0.25 – - 0.16]*	0.9003	75 ($8.02 \cdot 10^{-8}$)	0.42
Hyd_14	20	0.9208	100 ($8.69 \cdot 10^{-9}$)	0.80	[4.07 – 6.54]	[-4.58 – - 2.99]*	0.9073	84 ($3.48 \cdot 10^{-8}$)	0.86
Hyd_15	20	0.6649	14 ($1.38 \cdot 10^{-3}$)	1.21	[-1.99 – - 0.48]+	[0.17 – 0.61]+	0.5961	7 ($1.44 \cdot 10^{-2}$)	1.37
Hyd_16	20	0.9259	108 ($4.88 \cdot 10^{-9}$)	2.46	[8.71 – 13.39]*	[1.48 – 2.22]*	0.8935	69 ($4.91 \cdot 10^{-8}$)	2.97

Hyd_17	20	0.9182	97 ($1.15\cdot10^{-8}$)	0.52	[3.63 – 5.65]*	[-2.62 – - 1.70]*	0.8984	75 ($7.94\cdot10^{-8}$)	0.58
Hyd_18	20	0.8814	63 ($2.84\cdot10^{-7}$)	0.76	[13.98 – 15.13]*	[17.22 – 29.65]*	0.8546	49 ($1.65\cdot10^{-6}$)	0.84
Hyd_19	20	0.8832	65 ($2.50\cdot10^{-7}$)	0.50	[-5.65 – - 3.06]*	[4.38 – 7.50]*	0.8611	51 ($1.13\cdot10^{-6}$)	0.54
Hyd_20	20	0.8901	69 ($1.48\cdot10^{-7}$)	0.24	[1.25 – 1.61]*	[-3.42 – - 2.04]*	0.8545	48 ($1.78\cdot10^{-6}$)	0.28
Hyd_21	20	0.8163	36 ($1.14\cdot10^{-5}$)	2.19	[4.66 – 8.44]*	[-37.53 – - 18.06]*	0.7740	27 ($6.50\cdot10^{-5}$)	2.41
Hyd_22	20	0.8661	54 ($7.99\cdot10^{-7}$)	0.66	[0.97 – 1.96]*	[-8.45 – - 4.69]*	0.8344	41 ($4.89\cdot10^{-6}$)	0.73
Hyd_23	20	0.9046	81 ($4.40\cdot10^{-8}$)	1.07	[-36.23 – - 23.23]*	[-14.76 – - 9.17]*	0.8819	63 ($2.85\cdot10^{-7}$)	1.18
Hyd_24	19	0.9504	159 ($4.77\cdot10^{-10}$)	16.49	[73.60 – 98.50]*	[702.55 – 985.21]*	0.9382	125 ($3.00\cdot10^{-9}$)	18.37

* $p < 0.05$; + $p > 0.05$; Abb. = abbreviation of hydrophobicity scale; n = sample size; r = correlation coefficient;
F = Fisher parameter and associated type I error values (p); s = standard error of estimate; $_{95\%}$CI = 95% confidence interval;
Intercept = the intercept of the regression model; $p_{t\text{-}Stat}$ = the type I error for the intercept and slope on regression model (Student t test);
r_{loo} = correlation coefficient by leave-one-out analysis; F_{loo} = Fisher parameter by leave one out analysis;
s_{loo} = standard error of estimate by leave-one-out analysis

Table VI. Non-standard amino acids: predicted hydrophobicity

Abb.	Non-standard amino acid										
	Aib	Ciu	Dhd	Dop	Gab	Hcy	Hyp	Lth	Oth	Pyl	Sec
Hyd_01	0.71	-0.07	0.51	0.79	0.50	0.54	0.33	-0.21	0.38	0.25	0.18
Hyd_02	-9.91	-19.45	-11.39	-14.31	-10.67	-11.47	-11.88	-23.87	-13.92	-33.21	-12.87
Hyd_03	-0.22	-0.81	-0.16	5.04	0.14	0.06	-0.42	-1.04	0.07	-0.78	-1.14
Hyd_04	0.14	2.95	0.78	1.09	0.22	-0.29	0.20	3.19	2.35	2.98	10.97
Hyd_05	143.07	105.51	232.95	98.01	93.32	96.35	159.64	353.92	98.86	119.29	145.69
Hyd_06	-1.20	2.98	-0.23	-0.99	-1.07	-1.83	-1.66	3.33	2.09	0.63	5.19
Hyd_07	1.56	-1.57	0.22	-0.18	1.33	4.30	1.38	-1.69	-1.23	-1.59	-3.68
Hyd_08	0.68	-2.50	0.72	0.15	0.16	0.39	-0.07	-3.54	-0.61	-4.71	-2.03
Hyd_09	0.99	0.91	1.18	0.90	0.89	0.89	1.03	1.44	0.90	0.94	1.00
Hyd_10	0.45	-0.83	-0.07	0.17	0.46	0.49	0.49	-1.32	-1.33	-0.72	-13.33
Hyd_11	1.69	-8.25	-0.30	1.38	1.67	1.85	1.84	-15.63	-5.84	-5.70	-67.94
Hyd_12	0.46	-0.73	0.21	3.94	0.85	0.84	-0.09	-1.31	0.31	-2.42	-1.32
Hyd_13	-0.27	0.56	0.64	0.07	0.41	0.22	0.70	0.60	0.40	0.53	0.58
Hyd_14	0.90	-4.59	-0.13	1.25	-0.18	0.03	-1.23	-6.18	-0.92	-1.78	-2.37
Hyd_15	3.58	-1.02	-0.80	-0.97	3.27	-1.20	-1.10	-1.14	-0.53	-1.17	-0.59
Hyd_16	-1.77	6.16	4.04	1.36	4.53	1.12	1.63	7.63	5.57	17.39	10.79
Hyd_17	0.94	-2.38	0.73	1.19	0.33	0.44	-0.37	-2.89	-0.35	-2.05	-0.87
Hyd_18	16.31	15.23	18.89	15.02	14.89	14.97	16.79	22.35	15.04	15.63	16.39
Hyd_19	0.63	-0.74	0.38	2.23	0.46	1.10	1.63	-0.57	0.34	-3.58	-2.86
Hyd_20	1.22	0.15	0.98	0.86	1.19	1.39	1.20	0.06	0.38	0.14	-2.90
Hyd_21	-2.82	0.38	-2.76	5.66	0.14	5.78	0.53	4.56	0.11	1.25	5.18
Hyd_22	0.96	-1.61	0.38	0.09	0.89	1.36	0.91	-1.83	-1.06	-1.64	-8.96
Hyd_23	-1.87	0.88	-1.33	-5.41	0.00	-1.48	-3.27	-0.17	-0.96	-2.66	-0.39
Hyd_24	149.34	110.59	242.07	102.85	98.02	101.15	166.44	366.88	103.74	124.81	152.04

Discussions

Twenty-four hydrophobicity scales of standard amino acids were investigated by using the Molecular Descriptor family on the Structure-Property Relationship approach. A linear monovariate regression model was obtained for each hydrophobicity scale and the best model in terms of goodness-of-fit was analyzed. All regression models were statistically significant at a level of 5% (see Fisher parameter and significance, Table V). The power of determination, expressed as the determination coefficient, varied from 44% to 88% within the set of twenty standard amino acids. The lowest performance was obtained for Welling *et al.* scale. The hydrophobicity proved to be weakly related to atomic charge through topological interaction, according to this model (see Table III, sample abbreviated as Hyd_15). The

highest estimation power was obtained by the model on the Black *et al.* scale; 88% in the variation of hydrophobicity proved to be linearly related to the *lAmrLQg* molecular descriptor. This model shows that hydrophobicity is strongly related to atomic charge through geometric interaction.

The analysis of the character distribution in the descriptors name revealed that not all possible characters are found in the descriptors name of the best performing models (the proportion varied from 25% - for the fifth letter to 100% - for the seventh letter). Seventy-nine percent of the investigated scales showed that the hydrophobicity is related (in various degree) to atomic charges through geometric interaction. This observation supports the existence of a relationship between amino acids structure and their hydrophobicity and the similarities of these scales as well.

The empirical analysis of a correlation coefficient (Colton rules [49]) showed that, with a single exception (the model for hydrophobicity on the Welling *et al.* scale), very good correlation were obtained between the measured hydrophobicity and that estimated by the MDF SPR models.

A previously reported investigation of the ability of MDF SPR approach in modelling the amino acids hydrophobicity showed similar results. A set of fifteen standard amino acids was investigated on two scales and it was revealed that the hydrophobicity is strongly related to atomic charge through geometry interaction [50]. The comparison of the correlation coefficient of previously reported models ($p = 0.5244$ for Hessa *et al.* scale, and $p = 0.2586$ for Kyte & Doolitle scale) and that obtained in the present study showed no statistically significant difference.

The internal validation analysis proved the stability and validity of the models. The difference between the correlation coefficient obtained by the regression model in estimation vs prediction (leave-one-out analysis) varied from 0.01 to 0.07, the highest difference being obtained for the Welling *et al.* scale.

In conclusion, it can be said that the MDF SPR monovariate models provided good ability in investigated hydrophobicity of standard amino acids.

The hydrophobicity of each investigated non-standard amino acids was predicted based on the MDF SPR models derived on standard amino acids (Table VI). The impossibility to assess the reliability of the predicted values is the main limitation of the study. Due to limited resources, the measured hydrophobicity of non-standard amino acids, on

the investigated scales, was not available. The reliability analysis could be easily done once the hydrophobicity of non-standard amino acids is measured on each of twenty-four hydrophobicity scales.

A question arises during the investigation of the relationships between amino acids structures and their hydrophobicity by using the MDF SPR approach: "Is it possible to rescale the hydrophobicity scales?" This could be investigated by taking into consideration the scale that give the minimum value for the most standard amino acids (Sereda *et al.*, 1994), the scale that provide the maximum value for most amino acids (Manavalan & Ponnuswamy, 1978), a middle scale, and considering the MDF SPR models derived in the present investigation (the confidence values for intercept and slop associated to the regression models). This will be investigated in further research.

Acknowledgement

The research was supported by the National University Research Council Romania through grants (AT/93GR, ID_458, ID_1051).

References

[1] M. Szymański, J. Barciszewski, The genetic code--40 years on, *Acta Biochim. Pol.* **54** (2007) 51-54.

[2] T. Yoshimoto, Biochemistry and structural biology of microbial enzymes and their medical applications, *Yakugaku Zasshi* **27** (2007) 1035-1045.

[3] C. Cunchillos, G. Lecointre, Ordering events of biochemical evolution, *Biochimie* **89** (2007) 555-573.

[4] L.-E. Trudeau, R. Gutiérrez, On cotransmission & neurotransmitter phenotype plasticity, *Mol. Interv.* **7** (2007) 138-146.

[5] V. Demidchik, F.J.M. Maathuis, Physiological roles of nonselective cation channels in plants: From salt stress to signalling and development. *New Phytol.* **175** (2007) 387-404.

[6] T. Degenkolb, J. Kirschbaum, H. Brückner, New sequences, constituents, and producers of peptaibiotics: An updated review, *Chem. Biodivers.* **4** (2007) 1052-1067.

[7] R. Mazurkiewicz, A. Kuźnik, M. Grymel, A. Październiok-Holewa, α-Amino acid derivatives with a Cα-P bond in organic synthesis, *Arkivoc* **6** (2007) 193-216.

[8] P. Guan, I.A. Doytchinova, V.A. Walshe, P. Borrow, D.R. Flower, Analysis of Peptide-Protein Binding Using Amino Acid Descriptors: Prediction and Experimental

Verification for Human Histocompatibility Complex HLA-A*0201, *J. Med. Chem.* **48** (2005) 7418-7425.

[9] J.S. Nicolis, Super-selection rules modulating complexity: An overview, *Chaos, Solitons and Fractals* **24** (2005) 1159-1163.

[10] J.R. Cronin, S. Pizzarello, G.U. Yuen, Amino acids of the Murchison meteorite: II. Five carbon acyclic primary β-, γ-, and δ-amino alkanoic acids, *Geochim. Cosmochim. Acta* **40** (1985) 2259-2265.

[11] J. Liao, Protein and cellular engineering with unnatural amino acids, *Biotechnol. Prog.* **23** (2007) 28-31.

[12] D. Summerer, S. Chen, N. Wu, A. Deiters, J.W. Chin, P.G. Schultz, A genetically encoded fluorescent amino acid, *Proc. Natl. Acad. Sci. U.S.A.* **103** (2006) 9785-9789.

[13] C. Hansch, Quantitative approach to biochemical structure-activity relationships, *Acc. Chem. Res.* **2** (1969) 232-239.

[14] K. Roy, I. Sanyal, QSTR with extended topochemical atom indices. 7. QSAR of substituted benzenes to Saccharomyces cerevisiae. *QSAR Comb. Sci.* **25** (2006) 359-371.

[15] D. Haidacher, A. Vailaya, C. Horvath, Temperature effects in hydrophobic interaction chromatography, *Proc. Natl. Acad. Sci. U.S.A.* **93** (1996) 2290-2295.

[16] A. Rouhollahi, H. Shafieyan, J.B. Ghasemi, A QSPR study on the gc retention times of a series of fatty, dicarboxylic and amino acids by MLR and ANN, *Annali di Chimica* **97** (2007) 925-933.

[17] W. Ma, F. Luan, H. Zhang, X. Zhang, M. Liu, Z. Hu, B. Fan, Accurate quantitative structure-property relationship model of mobilities of peptides in capillary zone electrophoresis, *Analyst* **131** (2006) 1254-1260.

[18] Y. Lu, B. Bulka, M, DesJardins, S.J. Freeland, Amino acid quantitative structure property relationship database: A web-based platform for quantitative investigations of amino acids, *Protein Eng. Des. Sel.* **20** (2007) 347-351.

[19] S.D. Black, D.R. Mould, S.D. Black, D.R. Mould, Development of Hydrophobicity Parameters to Analyze Proteins Which Bear Post- or Cotranslational Modifications, *Anal. Biochem.* **193** (1991) 72-82.

[20] J. Kyte, R.F. Doolittle, A Simple Method for Displaying the Hydropathic Character of a Protein, *J. Mol. Biol.* **157** (1982) 105-132.

[21] W.C. Wimley, S.H. White, Experimentally determined hydrophobicity scale for proteins at membrane interfaces, *Nature Struct. Biol.* **3** (1996) 842-848.

[22] T. Hessa, H. Kim, K. Bihlmaier, C. Lundin, J. Boekel, H. Andersson, I. Nilsson, S.H. White, G. von Heijne, Recognition of transmembrane helices by the endoplasmic reticulum translocon, *Nature* **433** (2005) 377-381.

[23] J. Janin, Surface and inside volumes in globular proteins, *Nature* **277** (1979) 491-492.

[24] G.D. Rose, A.R. Geselowitz, G.J. Lesser, R.H. Lee, M.H. Zehfus, Hydrophobicity of amino acid residues in globular proteins, *Science* **229** (1985) 834-838.

[25] R. Wolfenden, L. Andersson, P. Cullis and C. Southgate, Affinities of Amino Acid Side Chains for Solvent Water, *Biochemistry* **20** (1981) 849-855.

[26] H.B. Bull, K. Breese, Surface tension of amino acid solutions: A hydrophobicity scale of the amino acid residues, *Arch. Biochem. Biophys.* **161** (1974â) 665-670.

[27] M.A. Roseman, Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds, *J. Mol. Biol.* **200** (1988) 513-522.

[28] G.W. Welling, W.J. Weijer, R. Van der Zee, S. Welling-Wester, Prediction of sequential antigenic regions in proteins, *FEBS Lett.* **188** (1985) 215-218.

[29] J.M.R. Parker, D. Guo, R.S. Hodges, New Hydrophilicity Scale Derived from High-Performance Liquid Chromatography Peptide Retention Data: Correlation of Predicted Surface Residues with Antigenicity and X-ray-Derived Accessible Sites?, *Biochemistry* **25** (1986) 5425-5431.

[30] T.J. Sereda, C.T. Mant, F.D. Sonnichsen, R.S. Hodges, Reversed-phase chromatography of synthetic amphipathic a-helical peptides as a model for ligand/receptor interactions effect of changing hydrophobic environment on the relative hydrophilicity/hydrophobicity of amino acid side- chains, *J. Chromatogr. A* **676** (1994) 139-153.

[31] R. Cowan, R.G. Whittaker, Hydrophobicity indices for amino acid residues as determined by HPLC, *Pept. Res.* **3** (1990) 75-80.

[32] T.P. Hopp, K.R. Woods, Prediction of protein antigenic determinants from amino acid sequences, *Proc. Natl. Acad. Sci. U.S.A.* **78** (1981) 3824-3828.

[33] P. Manavalan, P.K. Ponnuswamy, Hydrophobic character of amino acid residues in globular proteins, *Nature* **275** (1978) 673-674.

[34] J.L. Cornette, K.B. Cease, H. Margalit, J.L. Spouge, J.A. Berzofsky, C. DeLisi, Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. *J. Mol. Biol.* **195** (1987) 659-685.

[35] J.-L. Fauchere, V.E. Pliska, Hydrophobic parameters of amino-acid side-chains from the partitioning of N-acetyl-amino-acid amide, *Eur. J. Med. Chem.* **18** (1983) 369-375.

[36] D. Eisenberg, E. Schwarz, M. Komaromy, R. Wall, Analysis of membrane and surface protein sequences with the hydrophobic moment plot, *J. Mol. Biol.* **179** (1984) 125-142.

[37] M.J.K. Rao, P. Argos, A conformational preference parameter to predict helices in integral membrane proteins, *Biochim. Biophys. Acta.* **869** (1986) 197-214.

[38] K.J. Wilson, A. Honegger, R.P. Stotzel, G.J. Hughes, The behaviour of peptides on reverse-phase supports during high-pressure liquid chromatography, *Biochem. J.* **199** (1981) 31-41.

[39] R. Cowan, R.G. Whittaker, Hydrophobicity indices for amino acid residues asdetermined by HPLC, *Pept. Res.* **3** (1990) 75-80.

[40] D.M. Engelman, T.A. Steitz, A. Goldman, Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins, *Annu. Rev. Biophys. Biophys. Chem.* **15** (1986) 321-353.

[41] D.W. Urry, The change in Gibbs free energy for hydrophobic association - Derivation and evaluation by means of inverse temperature transitions, *Chem. Phys. Lett.* **399** (2004) 177-183.

[42] R.M. Sweet, D. Eisenberg, Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure, *J. Mol. Biol.* **171** (1983) 479-488.

[43] O.D. Monera, T.J. Sereda, N.E. Zhou, C.M. Kay, R.S. Hodges, Relationship of sidechain hydrophobicity and alpha-helical propensity on the stability of the single-stranded amphipathic alpha-helix, *J. Pept. Sci.* **1** (1995) 319-329.

[44] L. Jäntschi, Molecular Descriptors Family on Structure Activity Relationships 1. Review of the Methodology, *Leonardo Electr. J. Pract. Techn.* **6** (2005) 76-98.

[45] L. Jäntschi, S. Bolboacă, Results from the Use of Molecular Descriptors Family on Structure Property/Activity Relationships, *Int. J. Mol. Sci.* **8** (2007) 189-203.

[46] S. Bolboacă, L. Jäntschi, Amino Acids Sequences Analysis on Collagen, *B. Univ. Agricult.l Sci. Vet. Med. - Animal Sci. Biotech.* **63-64** (2007) 311-316.

[47] L. Jäntschi, S. Bolboacă, M.V. Diudea, Chromatographic Retention Times of Polychlorinated Biphenyls: from Structural Information to Property Characterization, *Int. J. Mol. Sci.* **8** (2007) 1125-1157.

[48] L. Jäntschi, S. Bolboacă, Structure-Activity Relationships on the Molecular Descriptors Family Project at the End, *Leonardo Electr. J. Pract. Techn.* **11** (2007) 163-180.

[49] T. Colton (Ed), *Statistics in Medicine*, Little, Brown, Boston, 1974; pp. 274.

[50] S. Bolboacă, L. Jäntschi, A Structural Informatics Study on Collagen, *Chem. Biol. Drug Des.* **71** (2008) 173-179.