

# IS AMINO ACIDS HYDROPHOBICITY A MATTER OF SCALE?

Sorana D. BOLBOACĂ<sup>1</sup> and Lorentz JÄNTSCHI<sup>2</sup>

<sup>1</sup> „Iuliu Hațieganu” University of Medicine and Pharmacy Cluj-Napoca, 13 Emil Isac, 400023 Cluj-Napoca, Romania, <http://sorana.academicdirect.ro>  
<sup>2</sup> Technical University of Cluj-Napoca, 15 Constantin Daicoviciu, 400020 Cluj-Napoca, Romania, <http://lori.academicdirect.org>

## ABSTRACT

Hydrophobic/hydrophilic character of an amino acid, an important property in protein structure and protein-protein interactions, is one of the most studied properties of amino-acids. To date, many hydrophobicity scales have been reported, the differences between them being significant: Janin (1979) and Kyte and Doolittle (1982) classify cistein as the most hydrophobic while Wolfenden et al. (1981) or Rose et al. (1985) do not. The aim of the research was to perform the structural modelling of amino acids hydrophobicities in order to identify and characterize the relationships between the structure and the property, and to quantify these relationships on a series of twenty-four scales.

## DATA SETS AND METHODOLOGY

The sample 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) has been investigated by using the Molecular Descriptors Family on Structure-Property Relationships approach [1]. The property of interest was represented by hydrophobic or hydrophilic character measured on twenty-four different scales [2] (see Table I). The amino acids structural information has been used in order to generate and calculate the Molecular Descriptors Family [3]. The best performing models in terms of goodness-of-fit were collected and analyzed [4]. The resulted models have been used in order to predict the hydrophobicity of a sample of eleven amino acids (seleno-L-cysteine, pyrrolysine, lanthionine, 2-aminoisobutyric acid, dehydroalanine, gamma-aminobutyric acid, ornithine, citrulline, homocysteine, hydroxyproline, and dopamine)

Table I. Amino acid hydrophobicity scales

Scale - Set abb.	Scale - Set abb.
Black <i>et al.</i> , 1991 - Hyd_01	Bull-Breese, 1974 - Hyd_13
Kyte-Doolittle, 1982 - Hyd_02	Roseman, 1988 - Hyd_14
Wimley-White, 1996 - Hyd_03	Welling <i>et al.</i> , 1985 - Hyd_15
Hessa <i>et al.</i> , 2005 - Hyd_04	Parker <i>et al.</i> , 1986 - Hyd_16
Sereda <i>et al.</i> , 1994 - Hyd_05	Cowan-Whittaker, 1990 - Hyd_17
Hopp-Woods, 1981 - Hyd_06	Manavalan-Ponnuswamy, 1978 - Hyd_18
Cornette <i>et al.</i> , 1987 - Hyd_07	Fauchere-Pliska, 1983 - Hyd_19
Eisenberg <i>et al.</i> , 1984 - Hyd_08	Rao-Argos, 1986 - Hyd_20
Janin, 1979 - Hyd_09	Wilson <i>et al.</i> , 1981 - Hyd_21
Rose <i>et al.</i> , 1985 - Hyd_10	Cowan-Whittaker, 1990 - Hyd_22
Engelman <i>et al.</i> , 1986 - Hyd_11	Urry, 2004 - Hyd_23
Sweet-Eisenberg, 1983 - Hyd_12	Monera <i>et al.</i> , 1995 - Hyd_24

## RESULTS

The interrelation between hydrophobicity of essential amino acids and their structural information was investigated. One monovariate model was obtained for each hydrophobicity scale. The model and its characteristics for each hydrophobicity scale are:

Amino acid property	Welling <i>et al.</i> , 1985	Wilson <i>et al.</i> , 1981	Amino acid property	Cornette <i>et al.</i> , 1987	Wimley-White, 1996	Amino acid property	Hopp-Woods, 1981	Cowan-Whittaker, 1990	Amino acid property	Manavalan-Ponnuswamy, 1978	Fauchere-Pliska, 1983
MDF SPR Equation	$\hat{y} = -1.23 + x(0.39)$	$\hat{y} = 6.55 + x(-27.79)$	MDF SPR Equation	$\hat{y} = -2.88 + x(-1.73)$	$\hat{y} = -3.37 + x(7.35)$	MDF SPR Equation	$\hat{y} = -1.99 + x(10.63)$	$\hat{y} = 1.47 + x(-6.57)$	MDF SPR Equation	$\hat{y} = 14.55 + x(23.43)$	$\hat{y} = -4.36 + x(5.94)$
SPR Determination (%)	44	66	SPR Determination (%)	69	71	SPR Determination (%)	74	75	SPR Determination (%)	78	78
MDF Descriptor (x)	amMRLQI	immRoQg	MDF Descriptor (x)	LmDROQg	iBmrWQI	MDF Descriptor (x)	iMPRoQg	AmDROQg	MDF Descriptor (x)	iMMPQg	ibDRPQg
Dominant Atomic Property	Charge (Q)	Charge (Q)	Dominant Atomic Property	Charge (Q)	Charge (Q)	Dominant Atomic Property	Charge (Q)	Charge (Q)	Dominant Atomic Property	Charge (Q)	Charge (Q)
Interaction via	Bonds (topology)	Space (geometry)	Interaction via	Space (geometry)	Bonds (topology)	Interaction via	Space (geometry)	Space (geometry)	Interaction via	Space (geometry)	Space (geometry)
Interaction Model	Q-d	Q <sup>1</sup>	Interaction Model	Q	Q <sup>2</sup> /d	Interaction Model	Q <sup>1</sup>	Q	Interaction Model	Q <sup>2</sup>	Q <sup>2</sup>
Structure on Property Scale	Inversed	Inversed	Structure on Property Scale	Logarithmic	Inversed	Structure on Property Scale	Inversed	Absolute	Structure on Property Scale	Inversed	Inversed
n = 20 r = 0.6649; F = 14; s = 1.21 r <sub>100</sub> = 0.5961; F <sub>100</sub> = 7; s <sub>100</sub> = 1.37	n = 20 r = 0.8163; F = 36; s = 2.19 r <sub>100</sub> = 0.7740; F <sub>100</sub> = 27; s <sub>100</sub> = 2.41	n = 20 r = 0.8309; F = 40; s = 1.70 r <sub>100</sub> = 0.7936; F <sub>100</sub> = 30; s <sub>100</sub> = 1.87	n = 20 r = 0.8434; F = 44; s = 0.48 r <sub>100</sub> = 0.8009; F <sub>100</sub> = 32; s <sub>100</sub> = 0.54	n = 20 r = 0.8608; F = 52; s = 1.01 r <sub>100</sub> = 0.8288; F <sub>100</sub> = 39; s <sub>100</sub> = 1.11	n = 20 r = 0.8661; F = 54; s = 0.66 r <sub>100</sub> = 0.8344; F <sub>100</sub> = 41; s <sub>100</sub> = 0.73	n = 20 r = 0.8814; F = 63; s = 0.76 r <sub>100</sub> = 0.8546; F <sub>100</sub> = 49; s <sub>100</sub> = 0.84	n = 20 r = 0.8832; F = 65; s = 0.50 r <sub>100</sub> = 0.8611; F <sub>100</sub> = 51; s <sub>100</sub> = 0.54				
Amino acid property	Rao-Argos, 1986	Janin, 1979	Amino acid property	Roseman, 1988	Rose <i>et al.</i> , 1985	Amino acid property	Urry, 2004	Engelman <i>et al.</i> , 1986	Amino acid property	Eisenberg <i>et al.</i> , 1984	Cowan-Whittaker, 1990
MDF SPR Equation	$\hat{y} = 1.43 + x(-2.73)$	$\hat{y} = 0.86 + x(1.74)$	MDF SPR Equation	$\hat{y} = -3.36 + x(3.76)$	$\hat{y} = 0.48 + x(-137.72)$	MDF SPR Equation	$\hat{y} = -29.73 + x(-11.96)$	$\hat{y} = 1.85 + x(-753.09)$	MDF SPR Equation	$\hat{y} = 1.68 + x(-0.92)$	$\hat{y} = 4.64 + x(-2.16)$
SPR Determination (%)	79	81	SPR Determination (%)	81	81	SPR Determination (%)	82	83	SPR Determination (%)	83	84
MDF Descriptor (x)	AmDROQg	imMpqQg	MDF Descriptor (x)	iBDdwQg	iHPrFQI	MDF Descriptor (x)	iBDMkEt	iNPrWQg	MDF Descriptor (x)	iAMkQg	iBmKQg
Dominant Atomic Property	Charge (Q)	Charge (Q)	Dominant Atomic Property	Charge (Q)	Charge (Q)	Dominant Atomic Property	Electronegativity (E)	Charge (Q)	Dominant Atomic Property	Charge (Q)	Charge (Q)
Interaction via	Space (geometry)	Space (geometry)	Interaction via	Space (geometry)	Bonds (topology)	Interaction via	Bonds (topology)	Space (geometry)	Interaction via	Space (geometry)	Space (geometry)
Interaction Model	Q	Q <sup>2</sup>	Interaction Model	Q <sup>2</sup> /d	Q <sup>2</sup> /d <sup>2</sup>	Interaction Model	Q <sup>2</sup> /d <sup>1</sup>	Q <sup>2</sup> /d	Interaction Model	Q <sup>2</sup> -d	Q <sup>2</sup> -d
Structure on Property Scale	Proportional	Inversed	Structure on Property Scale	Inversed	Logarithmic	Structure on Property Scale	Inversed	Logarithmic	Structure on Property Scale	Logarithmic	Logarithmic
n = 20 r = 0.8901; F = 69; s = 0.24 r <sub>100</sub> = 0.8545; F <sub>100</sub> = 48; s <sub>100</sub> = 0.28	n = 20 r = 0.8974; F = 74; s = 0.05 r <sub>100</sub> = 0.8744; F <sub>100</sub> = 58; s <sub>100</sub> = 0.06	n = 20 r = 0.8986; F = 75; s = 0.45 r <sub>100</sub> = 0.8812; F <sub>100</sub> = 62; s <sub>100</sub> = 0.48	n = 20 r = 0.8997; F = 76; s = 0.32 r <sub>100</sub> = 0.8599; F <sub>100</sub> = 56; s <sub>100</sub> = 0.36	n = 20 r = 0.9047; F = 81; s = 1.07 r <sub>100</sub> = 0.8819; F <sub>100</sub> = 63; s <sub>100</sub> = 1.18	n = 20 r = 0.9116; F = 89; s = 2.07 r <sub>100</sub> = 0.8731; F <sub>100</sub> = 51; s <sub>100</sub> = 2.56	n = 20 r = 0.9128; F = 90; s = 0.42 r <sub>100</sub> = 0.8935; F <sub>100</sub> = 70; s <sub>100</sub> = 0.46	n = 20 r = 0.9182; F = 97; s = 0.52 r <sub>100</sub> = 0.8984; F <sub>100</sub> = 75; s <sub>100</sub> = 0.58				
Amino acid property	Roseman, 1988	Sereda <i>et al.</i> , 1994	Amino acid property	Hessa <i>et al.</i> , 2005	Bull-Breese, 1974	Amino acid property	Parker <i>et al.</i> , 1986	Kyte-Doolittle, 1982	Amino acid property	Black <i>et al.</i> , 1991	Monera <i>et al.</i> , 1995
MDF SPR Equation	$\hat{y} = 5.30 + x(-3.78)$	$\hat{y} = 81.72 + x(817.95)$	MDF SPR Equation	$\hat{y} = -0.41 + x(7.18)$	$\hat{y} = 1.36 + x(-0.20)$	MDF SPR Equation	$\hat{y} = 11.05 + x(1.85)$	$\hat{y} = -7.60 + x(-19.17)$	MDF SPR Equation	$\hat{y} = 0.86 + x(-0.96)$	$\hat{y} = 86.05 + x(843.88)$
SPR Determination (%)	85	85	SPR Determination (%)	85	85	SPR Determination (%)	86	87	SPR Determination (%)	88	90
MDF Descriptor (x)	iAmrLQg	imMpqQg	MDF Descriptor (x)	AmDROQg	iPmLQI	MDF Descriptor (x)	iMPROQg	iGPdLQg	MDF Descriptor (x)	iAmrLQg	imMpqQg
Dominant Atomic Property	Charge (Q)	Charge (Q)	Dominant Atomic Property	Charge (Q)	Charge (Q)	Dominant Atomic Property	Charge (Q)	Charge (Q)	Dominant Atomic Property	Charge (Q)	Charge (Q)
Interaction via	Space (geometry)	Space (geometry)	Interaction via	Space (geometry)	Bonds (topology)	Interaction via	Space (geometry)	Space (geometry)	Interaction via	Space (geometry)	Space (geometry)
Interaction Model	Q-d	Q <sup>2</sup>	Interaction Model	Q	Q-d	Interaction Model	Q	d-Q	Interaction Model	d-Q	Q <sup>2</sup>
Structure on Property Scale	Logarithmic	Inversed	Structure on Property Scale	Proportional	Inversed	Structure on Property Scale	Logarithmic	Inversed	Structure on Property Scale	Proportional	Inversed
n = 20 r = 0.9208; F = 100; s = 0.80 r <sub>100</sub> = 0.9073; F <sub>100</sub> = 84; s <sub>100</sub> = 0.86	n = 20 r = 0.9232; F = 104; s = 20.73 r <sub>100</sub> = 0.9082; F <sub>100</sub> = 85; s <sub>100</sub> = 22.58	n = 20 r = 0.9238; F = 105; s = 0.32 r <sub>100</sub> = 0.9018; F <sub>100</sub> = 78; s <sub>100</sub> = 0.58	n = 20 r = 0.9252; F = 107; s = 0.36 r <sub>100</sub> = 0.9003; F <sub>100</sub> = 75; s <sub>100</sub> = 0.42	n = 20 r = 0.9259; F = 108; s = 2.46 r <sub>100</sub> = 0.8935; F <sub>100</sub> = 69; s <sub>100</sub> = 2.97	n = 20 r = 0.9327; F = 120; s = 1.11 r <sub>100</sub> = 0.9226; F <sub>100</sub> = 103; s <sub>100</sub> = 1.18	n = 20 r = 0.9376; F = 131; s = 0.12 r <sub>100</sub> = 0.9263; F <sub>100</sub> = 109; s <sub>100</sub> = 0.13	n = 19 (-Proline) r = 0.9504; F = 159; s = 16.49 r <sub>100</sub> = 0.9382; F <sub>100</sub> = 125; s <sub>100</sub> = 18.37				

The summary of characters in molecular descriptors used by the above-presented models is presented in Table II. Based on the obtained models, the hydrophobicity of 11 non-standard amino acids was obtained and is presented in Table III.

Table II. Character distribution in descriptors name

1 <sup>st</sup> letter	2 <sup>nd</sup> letter	3 <sup>rd</sup> letter	4 <sup>th</sup> letter	5 <sup>th</sup> letter	6 <sup>th</sup> letter	7 <sup>th</sup> letter
Cha. f.						
A 4	A 3	D 7	d 1	F 1	E 1	g 19
i 16	B 5	m 11	m 2	K 3	Q 23	i 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 = absolute frequency

Table III. 11 non-standard amino acids: predicted hydrophobicity

	Hyd_01	Hyd_02	Hyd_03	Hyd_04	Hyd_05	Hyd_06	Hyd_07	Hyd_08	Hyd_09	Hyd_10	Hyd_11	Hyd_12	Hyd_13	Hyd_14	Hyd_15	Hyd_16	Hyd_17	Hyd_18	Hyd_19	Hyd_20	Hyd_21	Hyd_22	Hyd_23	Hyd_24
Aib	0.71	-9.91	-0.22	0.14	143.07	-1.2	1.56	0.68	0.99	0.45	1.69	0.46	-0.27	0.9	3.58	-1.77	0.94	16.31	0.63	1.22	-2.82	0.96	-1.87	149.34
Ciu	-0.07	-19.45	-0.81	2.95	105.51	2.98	-1.57	-2.5	0.91	-0.83	-8.25	-0.73	0.56	-4.59	-1.02	6.16	-2.38	15.23	-0.74	0.15	0.38	-1.61	0.88	110.59
Dhd	0.51	-11.39	-0.16	0.78	232.95	-0.23	0.22	0.72	1.18	-0.07	-0.3	0.21	0.64	-0.13	-0.8	4.04	0.73	18.89	0.38	0.98	-2.76	0.38	-1.33	242.07
Dop	0.79	-14.31	5.04	1.09	98.01	-0.99	-0.18	0.15	0.9	0.17	1.38	3.94	0.07	1.25	-0.97	1.36	1.19	15.02	2.23	0.86	5.66	0.09	-5.41	102.85
Gab	0.5	-10.67	0.14	0.22	93.32	-1.07	1.33	0.16	0.89	0.46	1.67	0.85	0.41	-0.18	3.27	4.53	0.33	14.89	0.46	1.19	0.14	0.89	0	98.02
Hcy	0.54	-11.47	0.06	-0.29	96.35	-1.83	4.3	0.39	0.89	0.49	1.85	0.84	0.22	0.03	-1.2	1.12	0.44	14.97	1.1	1.39	5.78	1.36	-1.48	101.15
Hyp	0.33	-11.88	-0.42	0.2	159.64	-1.66	1.38	-0.07	1.03	0.49	1.84	-0.09	0.7	-1.23	-1.1	1.63	-0.37	16.79	1.63	1.2	0.53	0.91	-3.27	166.44
Lth	-0.21	-23.87	-1.04	3.19	353.92	3.33	-1.69	-3.54	1.44	-1.32	-15.63	-1.31	0.6	-6.18	-1.14	7.63	-2.89	22.35	-0.57	0.06	4.56	-1.83	-0.17	366.88
Oth	0.38	-13.92	0.07	2.35	98.86	2.09	-1.23	-0.61	0.9	-1.33	-5.84	0.31	0.4	-0.92	-0.53	5.57	-0.35	15.04	0.34	0.38	0.11	-1.06	-0.96	103.74
Pyl	0.25	-33.21	-0.78	2.98	119.29	0.63	-1.59	-4.71	0.94	-0.72	-5.7	-2.42	0.53	-1.78	-1.17	17.39	-2.05	15.63	-3.58	0.14	1.25	-1.64	-2.66	124.81
Sec	0.18	-12.87	-1.14	10.97	145.69	5.19	-3.68	-2.03	1	-13.33	-67.94	-1.32	0.58	-2.37	-0.59	10.79	-0.87	16.39	-2.86	-2.9	5.18	-8.96	-0.39	152.04
Pro																								96.57

## CONCLUSION

The amino acids hydrophobicity is a property linear related with compounds structure, being on strong relationship with atomic charge through geometry interaction, this fact being shown by the large majority of the scales.

## QUESTIONS THAT ARISES

First compelling question that arises from the investigation of amino acids hydrophobicity scales is as follows: "How reliable the predicted values are?" To answer to this question, future research will require. A study must be conducted in order to measure the hydrophobicity of the sample of non-standard amino acids. Once the measurements are done, the experimental values could be compared with predicted values.

Second compelling question that arises from this investigation is: "Is it possible to rescale the hydrophobicity scales?" This task could be made by taking into consideration the scale which in most of the cases give the minimum values for hydrophobicity (Sereda *et al.*, 1994), the scale that give the maximum values in most of the cases (Manavalan-Ponnuswamy, 1978), a middle scale, and considering the QSPR models obtained in the present research and the confidence intervals for intercepts and slopes. This will require future research.

## REFERENCES

- Jäntschi L, Bolboacă S.D. Modeling the octanol-water partition coefficient of substituted phenols by the use of structure information. *Int. J. Quant. Chem.* 2007;107:1736-1744.
- Bolboacă SD, Jäntschi L. Modeling Analysis of Amino Acids Hydrophobicity. *ChemMod - Chemical Graph Theory and Molecular Modeling Workshop*, October 23-26, Cl