STUDIA UBB CHEMIA, LXII, 2,Tom I, 2017 (p. 19-31)
(RECOMMENDED CITATION)
DOI:10.24193/subbchem.2017.2.02

Dedicated to Professor Costel Sârbu on the Occasion of His 65th Anniversary

CLASSIFICATION OF AMINO ACIDS BY MULTIVARIATE DATA ANALYSIS, BASED ON THERMODYNAMIC AND STRUCTURAL CHARACTERISTICS

OSSI HOROVITZ[a], ROXANA-DIANA PAŞCA[a*]

ABSTRACT. Principal component analysis (PCA) and cluster analysis (CA) were applied to classify 20 natural amino acids. We selected 18 characteristics, properties available from literature, as a basis for the classification. The correlations between these characteristics and their classification were investigated, as well as the classification of the amino acids. The results are presented as score plots of the first 3 principal components and as dendrograms obtained by clustering analysis. The resulting classification is consistent with the chemical behavior of amino acids and their mutual substitution possibilities in peptides and proteins.

Keywords: *principal component analysis, cluster analysis, amino acids, thermodynamic characteristics, structural characteristics*

INTRODUCTION

Amino acids, as building blocks for peptides and proteins were intensively studied and their importance in human nutrition and in animal feed, in food industry, pharmaceutical and cosmetics industries, as chelating agents, etc. The standard 20 amino acids, implied in the formation of peptides and proteins

[a] *Babeş-Bolyai University, Faculty of Chemistry and Chemical Engineering, 11 Arany Janos str., RO-400028, Cluj-Napoca, Romania*

** Corresponding author: rpasca@chem.ubbcluj.ro*

(proteinogenic amino acids) are enumerated here, together with their 3-letter and 1-letter symbols: glycine (Gly, G); alanine (Ala, A); phenylalanine (Phe, F); valine (Val, V); leucine (Leu, L); isoleucine (Ile, I); aspartic acid (Asp, D); glutamic acid (Glu, Q); asparagine (Asn, N); glutamine (Gln, E); serine (Ser, S); threonine (Thr, T); tyrosine (Tyr, Y); cysteine (Cys. C); methionine (Met, M); lysine (Lys, K); arginine (Arg, R); proline (Pro, P); histidine (His, H); tryptophan (Trp, W).

There are many possibilities to classify amino acids, according to the criteria selected to this aim. Based on their chemical structure [1] we can distinguish monoamino carboxylic acids (Gly, Ala, Val, Leu, Ile, Phe), monoamino dicarboxylic acids and their amides (Asp, Asn, Glu, Gln), hydroxyl amino acids (Ser, Thr, Tyr), thioamino acids (Cys, Met), diamino carboxylic acids and derivatives (Lys, Arg), heterocyclic amino acids (Pro, His, Trp). A classification based on structure and physical and chemical properties [2] groups the amino acids in: acidic (Asp, Glu), basic (Lys, Arg, His), aromatic (Tyr, Trp, Phe), S containing (Cys, Met), uncharged, hydrophilic (Ser, Thr, Asn, Gln), inactive hydrophobic (Gly, Ala, Val, Leu, Ile), special structure (Pro). The same amino acid can be assigned to more different classification groups based on the property considered [3]: polar / hydrophilic (Asn, Gln, Ser, Thr, Lys, Arg, His, Asp, Glu, [Cys, Tyr]); non polar / hydrophobic ([Gly], Ala, Val, Leu, Ile, Pro, Tyr, Phe, Trp, Met, Cys); forming hydrogen bond (Cys, Trp, Asp, Gln, Ser, Thr, Tyr, Lys, Arg, His, Asp, Glu); S containing (Cys, Met); negatively charged at neutral pH / acidic (Asp, Glu [Cys]; positively charged at neutral pH / basic (Lys, Arg, [His]; ionisable (Asp, Glu, His, Cys, Tyr, Lys, Arg); aromatic (Phe, Trp, Tyr, [His]; aliphatic (Gly, Ala, Val, Leu, Ile, Pro); forming covalent cross-bonding (S-S) (Cys); cyclic (Pro).

In view of the overlapping between the different classes, a Venn diagram showing the classification is helpful (Fig. 1) [4-6] This tries to group the amino acids according to their nature (aliphatic, aromatic), to the size of the molecules (small, tiny), to the hydrophobicity and in relation to their polarity (polar, charged - positive or negative).

Since the role of amino acids in the formation of proteins is of prime importance, a classification was proposed according to the interchangeability of different amino acids in the structure of a protein, without interfering with this structure [7]. The diagram in Figure 2 is a graphical representation of this substitutability. Amino acids connected in the diagram can be replaced with 95 % -probability. Marked with red are the solvent-exposed amino acids, with green – those located inside the protein molecule, according to their solvent exposed area (SEA)

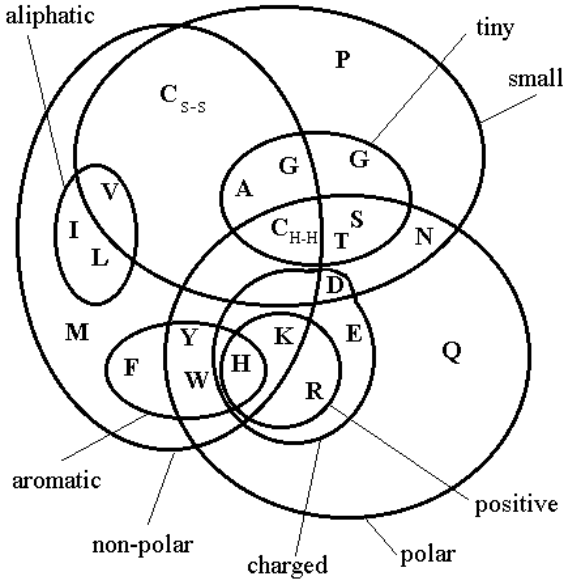

Figure 1. Venn diagram grouping the amino acids according to their properties (adapted from [4,5])

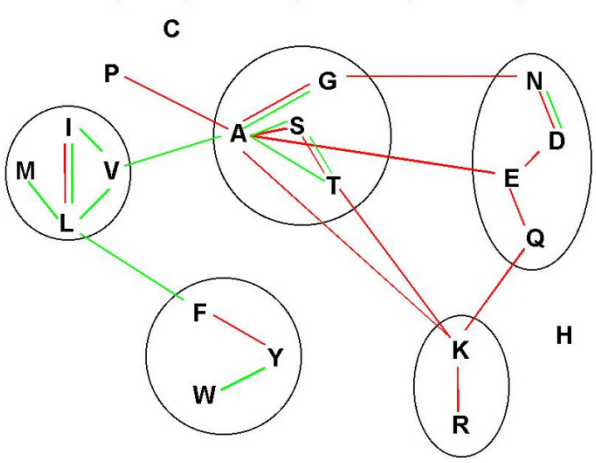

Figure 2. Possible substitutions of amino acids (according to [7]); amino acids bounded by lines have a substitution probability of 95%. Red: solvent exposed area (SEA) > 30 Å^2; green: SEA < 10 Å^2).

Here we propose a multivariate data analysis [8-14] of the amino acids, by principal components analysis (PCA) and cluster analysis (CA). PCA helps to reduce the number of variables necessary to describe a system, by maintaining the maximum possible information. Using the new variables as coordinates, "similarity maps" can be drawn for the analyzed system. In CA the (dis)similarity between elements is measured as "distances" between points in the space defined by the variables. The elements are then grouped in classes (clusters) by different clustering methods.

THEORETICAL METHODS

The objects of the classification were the 20 proteinogenic amino acids, enumerated above. As variables (descriptors, characteristics) 18 properties were chosen, as follows:

- Molar mass (MM)
- Acid dissociation constants $pK = -\lg K_a$ [3, 15, 16], corresponding to the first dissociation step (carboxylic group). pK_1 and the second dissociation (ammonium group), pK_2. Only seven amino acids present also a third dissociation constant, pK_3, due to a supplementary group (side chain), which was not considered in the PCA, but used in the calculation of the isoelectric point.
- Isoelectric point (pI), the pH where the molecules have no net electrical charge (zwitterions), and present a minimum solubility in water. The value can be calculated as the mean of the pK values corresponding to the equilibria which include the uncharged species.

Some descriptors for the elementary composition of the molecules were used:

- Number of carbon atoms (NC), a measure of the length of the carbon chain, and so related to the hydrophobicity of the molecule
- Number of hydrogen atoms (NH)
- Number of nitrogen atoms (NN), a measure of the number of basic functions
- Number of oxygen atoms (NO), a measure of the number of acid groups
- Number of sulphur atoms (NS)

Chou-Fasman Parameters for predicting the secondary structures in proteins [17, 18]

- P(α) – the probability for helix conformations (PA);
- P(β) – the probability for β-strands conformations (PB);
- P(t) – the probability for turns (PT)

Thermodynamic data

- Standard enthalpy of formation, H_{298}^0 (HF) [19]. For 4 amino acids (Phe, Trp, Gln, Arg) no values were found. Therefore, we calculated theoretical values for all aminoacids (except Cys and Met to avoid the complications related to the parametrization for S atoms), on the restricted Hartree-Fock level, by the semi empirical SCF-MO method PM3 [20], using the Hyper Chem 7.5 Software [21]. The molecular geometries for the zwitterionic forms were optimized by the Polak-Ribiere (conjugate gradient) algorithm. Between experimental (exp) and calculated (calc) values the following linear correlation was found:

$$\Delta H_{298}^o(\text{exp})\ [\text{kJ/mol}] = (-264 \pm 27) + (1.10 \pm 0.07)\ \Delta H_{298}^o(\text{calc}),\ r = 0.979,\ n = 14$$

This correlation was used to estimate the H_{298}^0 values for the 4 amino acids, where experimental values were not available.

- Standard state accessibility (AS), is defined as the average surface area of the residue X in a tripeptide Gly-X-Gly [22, 23].

- Average accessible surface area in proteins (AA)

- Solubility in water (SO) at 25 ^oC [15, 24]

- Hydrophobicity index (HP) [3, 25]. There are many different hydrophobicity scales, 46 scales were evaluated [26] and a PCA analysis was made on 40 scales [27]. One of the most used is the KD-scale [28], whose values very similar values with those used here. High positive values denote a strong hydrophobicity, while hydrophilic amino acids present negative values. In proteins, hydrophobic amino acids will be more probably located inside, while hydrophilic ones will be rather in contact with the aqueous environment.

- Melting points (MP) [29] for amino acids are quite high, an evidence for their zwitterionic character. They are a measure for the intermolecular interactions in solid state and for their stability. Some amino acids are decomposed before melting, for them we used the decomposition temperature.

The values used for the 20 amino acids are given in Table 1.

Table 1. Properties of the amino acids (the symbols are given in text)

Amino-acid	M (g/mol)	pK1	pK2	pK3	pI	nC	nH	nN	nO	nS	P(α)	P(β)	P(t)	ΔH_{298}° (kJ/mol)	AS (Å^2)	AA (Å^2)	SO (g/100g)	HP	MP (°C)
GLY,G	75.07	2.21	9.15		5.68	2	5	1	2	0	57	75	156	-527.5	88.1	25.2	24.99	0.67	233
ALA,A	89.09	2.35	9.87		6.11	3	7	1	2	0	142	83	66	-560	118.1	31.5	16.65	1	297
PHE,F	165.19	2.58	9.24		5.91	9	11	1	2	0	113	138	60	-440	222.8	28.7	2.965	2.5	283
LEU,L	131.17	2.36	9.6		6.04	6	13	1	2	0	121	130	59	-646.8	193.1	29	2.426	2.2	294
ILE,I	131.17	2.32	9.76		5.94	6	13	1	2	0	108	160	47	-640.6	181	23	4.117	3.1	284
VAL,V	117.15	2.3	9.6		5.96	5	11	1	2	0	106	170	50	-628.9	164.5	23.5	8.85	2.3	315
SER,S	105.09	2.21	9.15		5.68	3	7	1	3	0	77	75	143	-732.7	129.8	44.2	5.023	-1.1	228
THR,T	119.12	2.15	9.12		5.64	4	9	1	3	0	83	119	96	-776.3	152.5	46	9.7	-0.75	256
TRP,W	204.22	2.38	9.39		5.89	11	12	2	2	0	108	137	98	-350	266.3	41.7	1.136	1.5	289
PRO,P	115.13	1.99	10.6		6.3	5	9	1	2	0	57	55	152	-507.1	146.8	53.7	162.3	-0.29	221
ASP,D	133.1	1.88	9.6	3.65	2.77	4	7	1	4	0	101	54	146	-973.3	158.7	60.9	0.778	-3	270
ASN,N	132.12	2.02	8.8		5.41	4	8	2	3	0	67	89	156	-789	165.5	62.2	3.53	-2.7	234
GLU,E	147.13	2.19	9.67	4.25	3.22	5	9	1	4	0	151	37	74	-1003	186.2	72.3	0.864	-2.6	248
GLN,Q	146.15	2.17	9.13		5.65	5	10	2	3	0	111	110	98	-770	193.2	74	2.5	-2.9	185
TYR,Y	181.19	2.2	9.11	10.07	5.66	9	11	1	3	0	69	147	114	-685.6	236.8	59.1	0.0453	0.08	342
CYS,C	121.16	1.71	8.33	10.78	5.02	3	7	1	2	1	70	119	119	-534.1	146.1	13.9	0.011	0.17	260
MET,M	149.21	2.28	9.21		5.74	5	11	1	2	1	145	105	60	-577.5	203.4	30.5	3.381	1.1	281
ARG,R	174.2	2.18	9.09	13.2	11.15	6	14	4	2	0	98	93	95	-570	256	93.8	15	-7.5	244
LYS,K	146.19	2.2	9.2	10.28	9.59	6	14	2	2	0	114	74	101	-678.7	225.8	110.3	150	-4.6	224
HIS,H	155.16	1.78	8.97	5.97	7.47	6	9	3	2	0	100	87	95	-441.8	202.5	46.7	4.19	-1.7	287

All calculations were executed using the *Statistica* software package on the data matrix (20 objects, 18 variables). After the scaling of variables (to a mean of 0 and a variance of 1), PCA and CA procedures were applied, both for the classification of properties and of amino acids.

RESULTS AND DISCUSSION

Table 2. Table of correlations for the properties of the amino acids (significant correlations at P = 0.95 are bolded)

	MM	PK1	PK2	PI	NC	NH	NN	NO	NS	PA	PB	PT	HF	AS	AA	SO	HP	MP
MM	1.00	0.15	-0.18	0.23	**0.88**	**0.65**	**0.47**	0.03	-0.02	0.23	0.29	-0.24	0.17	**0.96**	0.37	-0.23	-0.20	0.26
PK1		1.00	0.30	0.12	0.43	0.44	-0.21	-0.24	-0.29	0.44	0.41	**-0.57**	0.19	0.26	-0.13	-0.38	0.42	0.25
PK2			1.00	-0.12	0.07	0.11	-0.31	-0.02	-0.41	0.23	-0.18	-0.16	-0.04	-0.14	-0.02	-0.09	0.24	0.07
PI				1.00	0.20	**0.58**	**0.73**	**-0.61**	-0.13	-0.05	0.10	-0.13	**0.47**	0.43	0.44	0.20	-0.42	-0.13
NC					1.00	**0.67**	0.22	-0.18	-0.21	0.18	**0.48**	-0.34	0.38	**0.87**	0.13	-0.24	0.17	0.44
NH						1.00	0.38	-0.35	-0.11	0.37	0.46	**-0.58**	0.19	**0.81**	0.34	-0.04	-0.07	0.21
NN							1.00	-0.19	-0.19	-0.01	-0.11	0.06	0.23	**0.53**	**0.57**	-0.11	**-0.69**	-0.23
NO								1.00	-0.22	0.03	-0.42	0.32	**-0.88**	-0.11	0.35	-0.17	-0.37	-0.19
NS									1.00	0.09	0.08	-0.09	0.18	-0.05	-0.36	0.26	0.16	0.06
PA										1.00	-0.00	**-0.77**	-0.14	0.31	0.06	-0.34	0.07	0.22
PB											1.00	**-0.55**	0.39	0.33	**-0.47**	-0.14	**0.59**	**0.58**
PT												1.00	-0.16	-0.36	0.24	0.23	-0.40	**-0.49**
HF													1.00	0.21	-0.39	0.06	0.39	0.25
AS														1.00	0.43	-0.18	-0.22	0.25
AA															1.00	0.14	**-0.87**	-0.46
SO																1.00	-0.16	-0.32
HP																	1.00	**0.55**
MP																		1.00

The covariance matrix, identical with the correlation matrix, since data were normalized, is given in Table 2. It confirms that the properties are correlated, and the number of variables can be reduced by PCA.

The eigenvalues (EV) and eigenvectors corresponding to each principal component were calculated. In Table 3, the EV for the first 6 principal components (PC) are given, together with the loadings of the characteristics associated with each PC. The contribution of each PC is given as % from the total variance of the system. The first three PC cumulate about 66% from the total variance, and the first 6 PC over 88%. In Table 3, the values representing the maximum contribution of each property to a PC are bolded.

Table 3. Eigenvectors and eigenvalues (EV) for the first 6 principal components (PC). The maximum contribution (loading) of each variable is bolded

Variable	PC1	PC2	PC3	PC4	PC5	PC6
MM	**0.732772**	-0.418786	0.149100	-0.437709	-0.119962	-0.106764
PK1	**0.580684**	0.323887	0.318470	0.405112	-0.049469	0.050468
PK2	0.036689	-0.222711	0.402239	**0.660723**	-0.184703	-0.279568
PI	0.455128	**-0.524516**	-0.519732	0.409233	-0.084511	0.101819
NC	**0.843749**	-0.107567	0.132647	-0.205899	-0.346962	-0.207651
NH	**0.845821**	-0.278541	0.043770	0.149541	0.143806	-0.252191
NN	0.321116	**-0.779242**	-0.270120	0.029570	-0.005560	0.398847
NO	0.475639	-0.251914	**0.735451**	-0.363289	-0.075591	-0.074636
NS	-0.068992	0.249191	-0.407097	-0.479635	**0.619484**	-0.087982
PA	0.422426	0.097598	0.532400	0.150091	**0.639621**	0.150662
PB	**0.646027**	0.455109	-0.196461	-0.224951	-0.123477	-0.057888
PT	**-0.693531**	-0.356646	-0.225276	-0.146697	-0.485925	-0.042470
HF	0.486677	0.233157	**-0.707765**	0.147943	-0.189069	0.132121
AS	**0.830771**	-0.454054	0.100198	-0.257977	-0.016263	-0.116410
AA	0.018882	**-0.924642**	0.237951	0.140003	0.039411	-0.155298
SO	-0.287173	-0.147992	-0.506858	0.090974	0.219967	**-0.713778**
HP	0.225330	**0.942794**	-0.044783	-0.001836	-0.142445	-0.093475
MP	0.514421	**0.518994**	0.105280	-0.285245	-0.121981	0.053338
EV	5.246110	4.066386	2.565046	1.679726	1.370267	0.985620
% total variance	29.14506	22.59103	14.25025	9.33181	7.61260	5.47567
Cumu-lated (%)	29.1451	51.7361	65.9863	75.3182	82.9308	88.4064

Classification of properties

Some characteristics with the largest contribution in PC1 are the number of H and C atoms, the AS surface and the molar mass, which are strongly correlated as shown in Table 2. Altogether there are 7 properties with maximum contribution in PC1. In PC2 there are 5 such properties, the most important being hydrophobicity (HP) and accessible surface area (AA), also strongly correlated (negatively). The other PC, 3 to 6, contain each only one or two properties with large contributions. In PC3 there are the number of O atoms (NO) and the enthalpy of formation (HF), also strongly correlated (Table 2): amino acids with more O atoms have actually lower H_{298}^0 values, with Asp and Glu the most stable (Table 1).

Such correlations can be visualized by 2- or 3-D loading plots; as an example, the scatter plot for the first two PC is given in Figure 2. Here related properties are represented by nearby points. But negatively correlated properties appear quite far from one another.

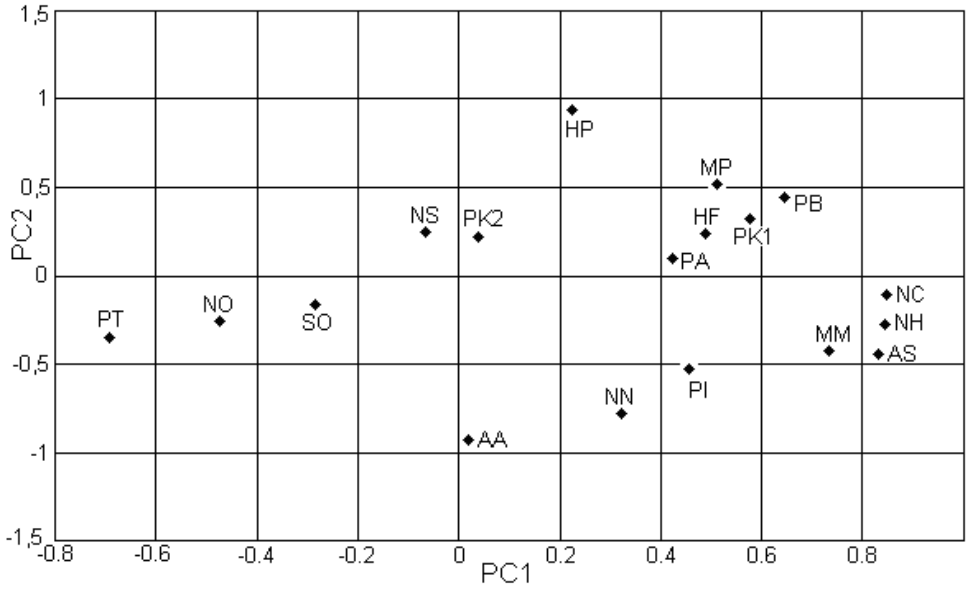

Figure 3. PCA loading plot for the two first principal components. The symbols of the characteristics correspond to those in Tables 1-3.

The results of cluster analysis can be presented as dendrograms, grouping the different properties in clusters. In Figure 4 is given such a diagram, obtained by the Complete Linkage method, using Euclidean distances

Classification of amino acids

In order to apply PCA for the classification of amino acids, we use the scores of each amino acid calculated for the first (more important) PCs. Here the first 3 PCs cumulate 66% of the total variance of data. Therefore, by reducing the number of 18 characteristics (natural variables) to 3 variables (the first 3 PCs) we retain enough information to characterize the system in 2D-representations (in the planes defined by PC1 and PC2, PC1 and PC3 or PC2and PC3) an 3D- representation (in the space of PC1, PC2 and PC3). As an example, in Figure 5 the representation in the PC1-PC2 plane is given. Some groupings of amino acids are highlighted in the figure.

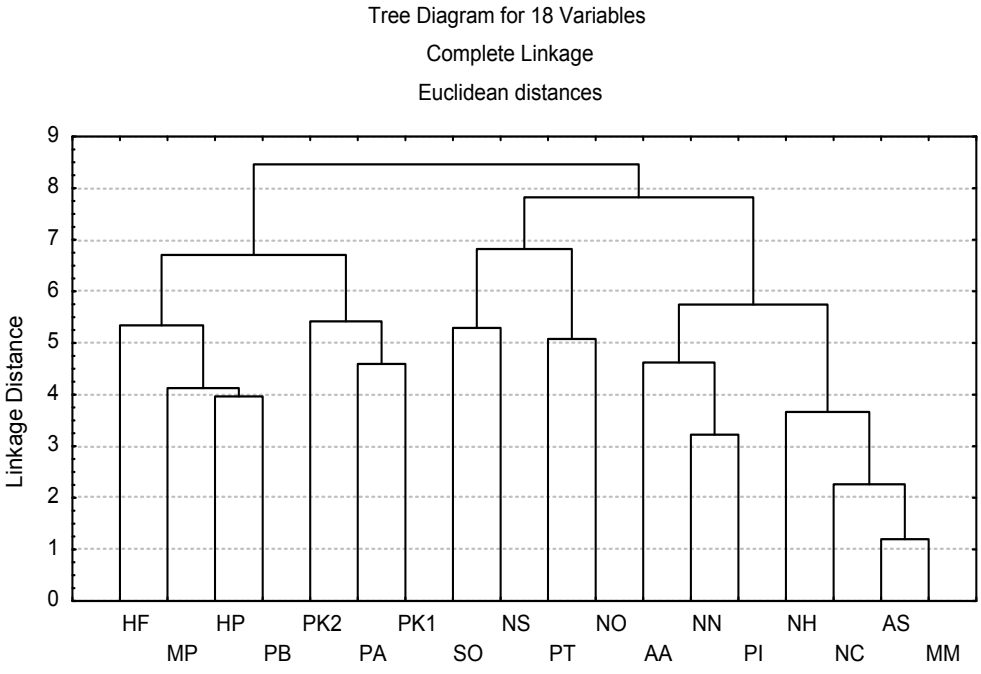

Figure 4. Dendrogram for hierarchical clustering for 18 properties

The similarities between amino acids can also be followed on the dendrogram (Figure 6) obtained in CA, using the same Complete Linkage method, with Euclidean distances, as for the properties.

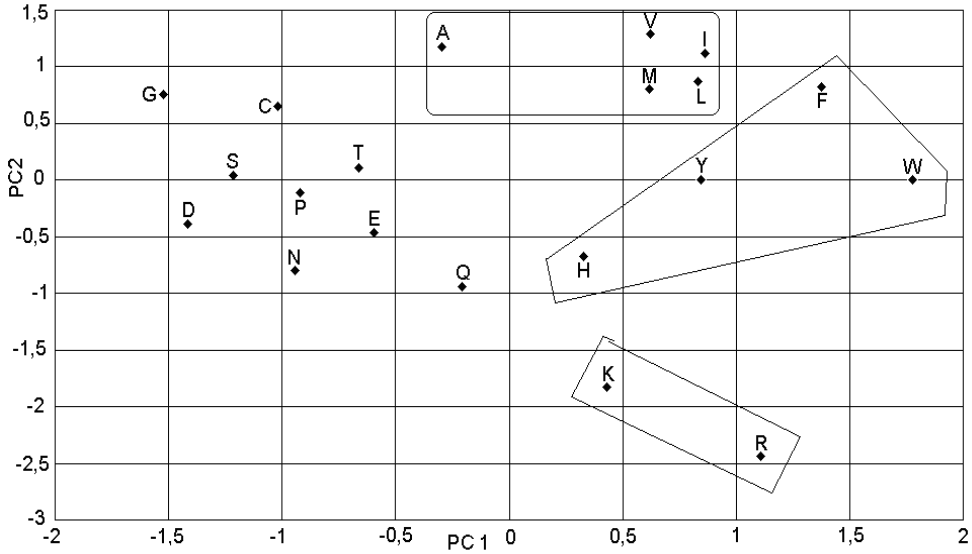

Figure 5. PCA score plots for the 20 amino acids in the plane of the two first principal components. 1-letter symbols are used for the amino acids.

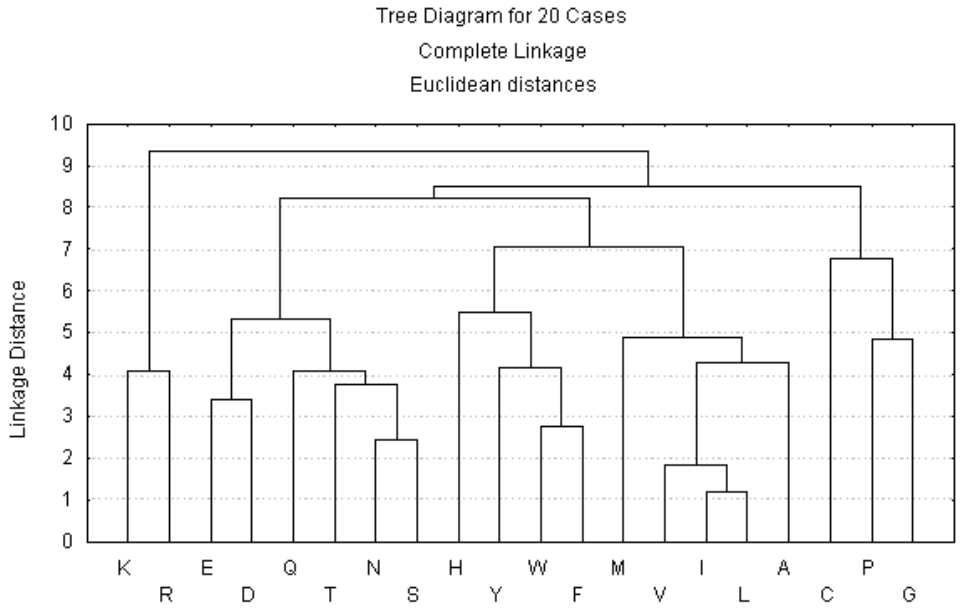

Figure 6. Dendrogram for hierarchical clustering by complete linkage method for 20 amino acids. 1-letter symbols are used for the amino acids (see Introduction)

The strongest related amino acids according to CA are leucine (L) and isoleucine (I), isomers with very similar properties. In the dendrogram they are joined by valine (V), methionine (M), and alanine (A). They all are aliphatic hydrophobic amino acids. Another cluster, more loosely connected, contains phenylalanine (F), tryptophan (W), tyrosine (Y), and histidine (H) - all aromatic amino acids, mostly hydrophobic, but also partially hydrophilic, particularly H. These two clusters are then connected (Fig. 6), giving a cluster of mostly hydrophobic amino acids.

Lysine (K) and arginine (R) form a cluster of basic amino acids, positively charged in proteins at physiological pH, histidine, also basic, is near to them in Fig. 5. The two acidic amino acids, negatively charged in proteins: aspartic acid (D), and glutamic acid (E), are also united in a cluster (Fig. 6). Asparagine (N) and serine (S) give a cluster with glutamine (Q), and threonin (T): they are neutral in proteins, but contain polar groups and show hydrophilic properties. Quite loosely interconnected are glycine (G), proline (P), and cysteine (C) – small molecules, rather dissimilar to other amino acids

It is worth noting that though the characteristics used in classification were mostly of physical and stoichiometrical nature, the resulted classification follows quite well the chemical properties and the character of amino acids. It is also interesting to compare the clustering results with the possibilities of mutual interchange of amino acids in proteins [7] (Fig. 2). We find there the same group L, I, M, V as in CA, as well the group F, Y, W, or the group K, R. The other possibilities of substitution in Fig.2 are also quite well reflected in PCA and CA.

CONCLUSIONS

The use of the methods of multivariate analysis applied for the standard 20 proteinogenic amino acids, based on literature data for 18 structural and physico-chemical characteristics, resulted in a classification able to predict the chemical behavior of these compounds. It demonstrates the possibilities of principal component analysis and cluster analysis in the description of different classes of chemical compounds.

ACKNOWLEDGMENTS

We thank Prof. Dr. Costel Sârbu, who kindly initiated us in multivariate analysis.

REFERENCES

1. C.D. Neniţescu, "Chimie organică", Vol. II, 6th Edition, EDP Bucureşti, **1968**.
2. ***, „Molecular Biology Web Book", Chap. 2, „Protein Structure and Function", *http://www.web-books.com/MoBio/Free/Chap2.htm* (accessed September **2016**).
3. U.C. Davis, „Biological Sciences: Amino Acid Structure/Function", *https://www.mcb.ucdavis.edu/courses/bis102/AAProp.html* (accessed September **2016**).
4. W.R. Taylor, *Journal of Theoretical Biology,* **1986**, *119*, 205.
5. M.J. Betts, R.B. Russell. „Amino acid properties and consequences of substitutions". *In* „*Bioinformatics for Geneticists*", M.R. Barnes, I.C. Gray eds., Wiley, Chichester, **2003**.
6. O. Horovitz, A. Mocanu, Gh. Tomoaia, M. Crisan, L.-D. Bobos, Cs. Racz, M. Tomoaia-Cotisel*, Studia UBB Chemia*, **2007**, *52*, 53.
7. D. Bordo, P. Argos, *Journal of Molecular Biology,* **1991**, *217*, 721.
8. T.P.E. auf der Heyde, *Journal of Chemical Education,* **1990**, *67*, 461.
9. H. Pop, C. Sârbu, O. Horovitz, D. Dumitrescu, *Journal of Chemical Information and Computer Sciences,* **1996**, *36*, 465.
10. C. Sârbu, O. Horovitz, H.Pop, *Journal of Chemical Information and Computer Sciences,* **1996**, *36*, 1098.
11. O. Horovitz, C.Sarbu, H.F.Pop, „Clasificarea rațională a elementelor chimice", Ed. Dacia, Cluj-Napoca, **2000**.
12. O.Horovitz, C.Sârbu, H.F.Pop, *Revista de Chimie (Bucharest)*, **2000**, *51*, 17.
13. O. Horovitz, C. Sârbu, *Journal of Chemical Education*, **2005**, *82*, 473.
14. O. Horovitz, C. Sârbu, *Revista de Chimie (Bucharest)*, **2006**, *57*, 413.
15. J. Reichert, A. Jabs, P. Slickers, J. Sühnela, *Nucleic Acids Research*, **2000**, *28,* 246.
16. D.R. Lide, "Handbook of Chemistry and Physics", 12nd Ed., CRC Press, Boca Raton, **1991**.
17. P.Y. Chou, G.D. Fasman, *Biochemistry,* **1974**, *13*, 211.
18. P. Prevelige, Jr., G.D. Fasman, *Chou-Fasman, Prediction of Secondary Structure, in Prediction of Protein Structure and the Principles of Protein Conformation*, ed. G. B. Fasman, Plenum Press, New York, **1989**.
19. ***, Chemistry WebBook, NIST Standard Reference Database Number 69 - June, *http://webbook.nist.gov/chemistry/* (accessed September **2016**).
20. . J.J.P. Stewart, *Journal of Computational Chem*istry, **1989**, *10*, 209.
21. ***, HyperChem7.5 Molecular Modeling System, Hypercube Inc., Gainesville, **2002**.
22. C. Chotia, *J. Mol. Biol*., **1975**, *105*, 1.
23. ***, CRC Handbook of Chem.& Phys., 58 Ed., Cleveland, Ohio, **1977**.
24. D. Eisner, R.M. Weiss, T.C. Terwilliger, William Wilcox, *Faraday Symp. Chem. Soc.,* **1982**, *17*, 109.

25. J.L. Cornette, K.B. Cease, H. Margalit, J.L. Spouge, J.A. Berzofsky, C. DeLisi, *J. Mol. Biol.,* **1987**, *195*, 659.
26. A. Neumaier, W. Huyer, E. Bornberg-Bauer, „Hydrophobicity Analysis of Amino Acids" *http://www.mat.univie.ac.at/~neum/software/protein/aminoacids.html* (accessed September **2016**).
27. J. Kyte, R.F. Doolittle, *Journal of Molecular Biology,* **1982**, *157*, 105.
28. R.C Beavis, D. Fenyo, "Amino Acid Properties of Interest in Protein Chemistry*", http://prowl.rockefeller.edu/aainfo/contents.htm* (accessed September **2016**).