Babeş-Bolyai University
Faculty of Chemistry and Chemical Engineering
Specialization Chemistry

Alexandrina GUIDEA

EXTENDED SUMMARY

Modeling and determination of specific properties of amino acids and related compounds using advanced chemometric methods

Scientific advisor:
Prof. Emerit Costel SÂRBU

CHAPTER I—THEORETICAL CONSIDERATIONS 7
1. General considerations—Amino acids 7
1.1. Structure of amino acids 10
1.2. Amino acid chirality 11
1.3. Amino acid side chain 11
1.4. Clasification of amino acids 12
1.5. Essential, non-essential and conditional amino acids 14
1.6. Amino acid property 16
1.7. The role of amino acids 20
2. Chemometrical methods 24
2.1. Regression analysis 24
2.2. Principal component analysis 25
2.3. Cluster analysis 26
2.3.1. Grouping methods 27
2.3.2. Estimation of distances 27
2.4. Discriminant analysis 29
2.4.1. Canonica analysis 31
2.5. Sum of ranking differences 31
2.6. Analyzes based on fuzzy set theory 32
3. Thesis objectives 33
CHAPTER II— STUDY OF LITERATURE AND DEFINITION OF PREMISES 35
4. Specific properties of amino acids 35
4.1. Antioxidant activity 35
Antioxidant activity methods 37

4.2. Modeling and prediction of lipophilicity of amino acids 40
4.2.1 Building QSPR models 42
4.2.2 Molecular descriptors 44
CHAPTER III—ORIGINAL CONTRIBUTION 48
5. Determination of antioxidant activity and chelating capacity of amino acids 48
5.1. Introduction 48
5.2. Materials and methods 49
5.2.1. Samples 49
5.2.2. Sample preparation 49
5.2.3. Antioxidant activity methods 49
5.2.3.1. DPPH radical scavenging capacity assay, DPPH 50
5.2.3.2. ABTS radical cation decolorization assay, ABTS 51
5.2.3.3. Superoxide radical scavenging activity assay, SORS 52
5.2.3.4. Nitric oxide radical scavenging assay, NO 54
5.2.3.5. Ferric ions reducing antioxidant power, FRAP 55
5.2.3.6. Cupric ions reducing antioxidant power assay, CUPRAC 57
5.2.3.7. Chromium reducing antioxidant capacity assay, CHROMAC 59
5.2.3.8. Chelatiang Capacity Assay 60
5.2.4. Chemometric analysis 61
5.3. Results and discussion 62
5.3.1. Radical scavenging capacity 66
5.3.2. Reducing antioxidant power 69
5.3.3. Chelating capacity 72
5.3.4. Cluster analysis and principal component analysis 73
5.3.5. Results of SRD analysis 80
5.4. Conclusions 82
6. Determination of antioxidant activity and chelating capacity of some biogenic amines and related drugs 83

6.1. Introduction 83
6.2. Materials and methods 84
6.2.1 Samples 84
6.2.2. Antioxidant activity and chelating capacity assays 87
6.2.3. Chemometric analysis 87
6.3. Results and discussion 87
6.4. Conclusions 97
7. Modeling of amino acids lipophilicity using QSPR studies 98
7.1. Introduction 98
7.2. Materials and methods 101
7.2.1. Amino acids structure 101
7.2.2. Molecular descriptors 101
2D descriptors 103
3D descriptors 105
7.2.3. Selection of variables (descriptors) 107
7.2.4. Quality parameters 109
7.3. Results and discission 111
7.4. Conclusions 126
8. Fuzzy characterization and classification of solvents according to their polarity and selectivity. A comparison with the *Snyder* approach.. 127
8.1. Introduction 127
8.2. Materials and methods 128
8.2.1. Dataset 128
8.2.2. Fuzzy classification methods 128
8.2.2.1. Fuzzy classification 128
c-means fuzzy clustering 130

Fuzzy linear discriminant analysis 131
8.3. Results and discussion 132
8.3.1. Fuzzy classification using c-means fuzzy clustering 133
8.3.2. Fuzzy linear discriminant analysis 140
8.4. Conclusions 144
9. Characterization and classification of foods based on amino acids content, using fuzzy c-means algorithms 145
9.1. Introduction 145
9.2. Materials and methods 146
9.2.1. Samples 146
9.2.2. Methods 149
9.3. Results and discussion 149
9.4. Conclusion 154
CHAPTER IV—GENERAL CONCLUSIONS 155
BIBLIOGRAPHY 159
Appendix 1. Degrees of membership for each fuzzy partition obtained by applying fuzzy c-means algorithm. 176
Appendix 2. Degreees of membership for each fuzzy partition obtained by applying the fuzzy c-means algorithm. 178
Appendix 3. Canonical scores corresponding to all ROOTs. 180
Appendix 4. List of published papers. 182

KEYWORDS:

Amino acids

Antioxidant activity

Lipophilicity index

Quantitative Structure—Property/Activity Relationship (QSRR/QSPR/QSAR)

Reverse phase thin layer chromatography (RP-TLC)

Fuzzy classification

Principal component analysis

Cluster analysis

Discriminant analysis

Sum of ranking differences

1. Introduction

Amino acids (AA) are organic compounds that contain one or more amino groups ($–NH_2$) in the molecule along with one or more carboxyl groups (–COOH). AA can be considered derivatives of carboxylic acids in which one or more hydrogen atoms are substituted with $–NH_2$ groups. Many of AAs found in nature, only 20 of them (α-AA) are found as building blocks of proteins, being called proteinogenic AAs. Due to their high mass (representing 40–45% of body weight), skeletal muscle is the largest reservoir of proteinogenic AA in the whole body [1]. AAs that are not found in the composition of proteins and peptides are called non-proteinogenic AAs.

All α-AAs have in common the $–NH_2$ and –COOH groups attached on α carbon atom (Cα) and differ from each other by the side chain (R). The mode, order and number of AAs that make up each specific protein are written in the genetic code (DNA and RNA), where genes contain the hole "instructions" for their synthesis [2].

No group of compounds is more difficult to characterize than AA! These are basic structural elements, which are part of proteins, compounds with remarkable biological properties. They function as precursors for the synthesis of a wide range of biological substances, such as: hormones, purines, pyrimidines, porphyrins, vitamins, sugars, cofactors or biogenic amines [3,4]. It also plays an important role in cellular signaling, acting as regulators of gene expression, protein phosphorylation, nutrient transport (hemoglobin heme), cellular metabolism, and innate and cell-mediated immune responses [5].

The relationship between AA and proteins, as well as their specific properties in proteins, is the driving force behind various research on these fascinating compounds.

The present thesis aims to determine, characterize and classify AAs and related compounds in terms of lipophilicity and antioxidant activity using various classical and advanced chemometric methods. The paper is structured in two parts, a theoretical part and an experimental part.

The theoretical part is structured in four chapters, that deals with aspects of chemometric methods used to characterize and classify AAs and related compounds, experimental methods for determining antioxidant activity and aspects of modeling and predicting of lipophilicity of AAs. The information presented in this section is supported by many bibliographic references from the last ten years.

The experimental part includes five chapters dedicated to the determination of the antioxidant activity of AAs and biogenic amines using differents analytical methods, modeling and prediction of lipophilicity (retention coefficient R_{M0}) of AAs, as well as classification of solvents involved in chromatographic determination of lipophilicity and food classification in terms of AAs content applying classical and advanced chemometric methods based on fuzzy set theory. The results of the research included in this part have been presented and published in prestigious journals abroad and in the country.

CHAPTER II—STUDY OF LITERATURE AND DEFINITION OF PREMISES

2. Specific properties of amino acids

2.1. Antioxidant activity

The oxidation products of AAs differ depending on the structure from which they are derived. Following oxidation, some AAs can be hydroxylated, nitrated, nitrosylated and even converted to a carbonyl derivative by extraction of a hydrogen atom from the parent molecule. Biogenic amines derived from AA tyrosine are called catecholamines because they contain a catechol or 3,4-dihydroxylphenyl group. The most abundant catecholamines are epinephrine (adrenaline), norepinephrine (noradrenaline) and dopamine. Catecholamines are hormones released in the body in response to a series of stresses to regulate the body's physiological functions [6–10]. In addition to these physiological functions, they can also act as antioxidants or prooxidants [11,12].

All these radical-induced oxidation products have a high functional importance because they contribute to the modulation of protein activity *in vivo*. Moreover, some AAs, such as aromatics and acidics, act as biomarkers of oxidative stress under various conditions or as indicators of the effectiveness of a treatment (by increasing or decreasing their level in the blood) [13].

The role of AAs as targets of various radical attacks, as well as their role in antioxidant defense, have led to the fact that AAs can be considered sacrificial (or primary) antioxidants with the ability to inhibit free radicals [14].

Till now, no extensive reports have been published in the literature regarding the antioxidant activity of free proteinogenic AAs. Previous studies [15–17] show the antioxidant activity of small groups of AAs and compare their activity with reference antioxidants such as ascorbic acid and Trolox. But the results obtained are based only on a few antioxidant activity methods. A comprehensive evaluation has not yet been provided to better understand the antioxidant behavior of these important compounds.

Recent studies [18,19] have emphasized that peptides have a substantial antioxidant activity and not proteins. Glutathione, the most important endogenous antioxidant of the cell, is a tripeptide consisting of 3 AAs containing cysteine, being also an indicator of the level of oxidative stress expressed by the ratio between reduced glutathione and oxidized.

While hydrolyzed proteins have important antioxidant activity, it is not yet possible to explain exactly how AAs in their composition influences their ability to inhibit lipid peroxidation [20]. Previous studies [21] note that there are several specific AAs, mainly basic (histidine and lysine) and tyrosine, which modulate the antioxidant activity of peptides. Thus, understanding the relationship between the AAs composition of peptides and their antioxidant activity could lead to the development of a new class of natural multifunctional antioxidants. From this point of view, AAs are unique compared to other antioxidants because, in addition to the effects they have on the body, they can be used as potential antioxidants to inhibit various oxidation processes.

2.2. Modeling and prediction of lipophilicity of amino acids

Computational chemistry is currently an efficient and elegant way to organize the properties of compounds, but also to develop strategies for developing and synthesizing new structures based on a molecular design.

Quantitative structure-property (QSPR) or –activity (QSAR) relationship represents the mathematical correlation between the specific molecular property or biological activity and one or more physico-chemical and/or structural molecular characteristics, known as descriptors.

QSPR studies have led to increased drug discovery efforts to a large extent solely on the basis of theoretical information obtained for the compounds studied [22–24]. The reasons for the widespread use of QSPR studies are mainly the following: (a) to reduce the time and cost of design studies; (b) to predict biological, pharmaceutical, physical and chemical properties; (c) to assist practitioners by providing valuable information found in large databases; (d) to understand the mechanism of action of the property of interest. Therefore, QSPR studies have found wide applications in life sciences [25] (biology, agriculture and medicine) as well as in physics, chemistry and engineering [26] (organic chemistry, physical chemistry, materials science).

Despite widespread use, the full potential of QSPR models has not yet been achieved, with efforts being made to generate models with high predictive performance. To benefit, QSPR models must be simple and easy to interpret so that the most important features can be easily highlighted.

A robust, reliable and reproducible QSPR model can be obtained by selecting those descriptors that strongly correlate with the property of interest and by a very careful analysis of the data by an expert and well-trained staff. Not every set of descriptors can lead to promising results.

Lipophilicity (hydrophobicity) is a fundamental molecular property defined as the logarithm of the octanol-water partition coefficient ($logP_{OW}$) and which practically reflects the partition of the non-ionized compound between two phases, usually octanol and water [27,28].

Various experimental (chromatographic) methods have also been applied and continue to be used successfully to estimate the physico-chemical characteristics of chemical compounds, of which lipophilicity is one of the most important [29].

Modeling the specific properties (lipophilicity and antioxidant activity) of proteinogenic AA as well as predicting these parameters using different experimental and calculated molecular descriptors using advanced chemometric methods will allow a better understanding of the relationships between the structure of AAs and their physicochemical and biochemical properties. At the same time, the new information provided in the QSPR models can provide a more in-depth (comprehensive) characterization of AAs.

CHAPTER III—ORIGINAL CONTRIBUTIONS

3. Determination of antioxidant activity and chelating capacity of amino acids

3.1. Introduction

The purpose of this chapter is to evaluate the antioxidant power of proteinogenic AAs, such as free radical scavenging and free radical scavenging activity and also, metal chelation capacity. At the same time, the comparison of the results obtained using classical and advanced chemometric methods such as hierarchical cluster analysis (HCA), PCA and SRD.

3.2. Materials and methods

3.2.1. Samples

A set of twenty proteinogenic amino acids were used in this study. They includes: essential AA (EAA, **red**), non-essential AA (NEAA, **blue**) and conditional essential AA (CEAA, **green**): Alanine (**Ala**), Arginine (**Arg**), Asparagine (**Asn**), Aaspartic acid (**Asp**), Cysteine (**Cys**), Glutamine (**Gln**), Glutamic acid (**Glu**), Glycine (**Gly**), Hystidine (**His**), Isoleucine (**Ile**), Leucine (**Leu**), Lysine (**Lys**), Methionine (**Met**), Phenilalanine (**Phe**), Proline (**Pro**), Serine (**Ser**), Threonine (**Thr**), Triptophan (**Trp**), Tyrozine (**Tyr**) și Valine (**Val**) with high analitical grade (98–99%) obtained from Merck and Sigma (Sigma-Aldrich GmbH, Sternheim, Germany).

3.2.2. Samples preparation

The standard solutions of all amino acids were prepared in ultrapure water at concentration 0.1 M. For those amino acids that were not hydrochlorides, stoichiometric hydrochloric acid was added.The final concentration of each amino acid in all the employed assays was 5 mM for ABTS, DPPH, SORS, FRAP, CUPRAC assay and metal chelating capacity, 20 mM for NO scavenging assay and 17 mM for CHROMAC assay.

3.2.3. Methods of determination of antioxidant activity

For derermination of antioxidant activity were used eight methods groups as radical scavenging assays (DPPH, ABTS, SORS şi NO), reducing antioxidant power assays (FRAP, CUPRAC şi CHROMAC) and metal chelating capacity.

3.2.4. Chemometric methods

For realization of purpose, were used the following chemometric methods for multidimensional data analysis, such as HCA, PCA and SRD analysis.

3.3. Results and discussion

3.3.1. Radical scavenging capacity

In **Figure 3.1** it can be seen that some of the AAs that contain mainly heteroatoms (Cys, Glu, Asp) or aromatic fragments (Tyr, Phe, Trp) in the side chain, have a significant ability to quench DPPH radicals. With the exception of Cys and Arg, which have a much stronger antioxidant reactivity than the ABTS radical compared to DPPH, the other AAs investigated behave differently or at the limit.

At the same concentration, the following amino acids showed the highest superoxide radical quenching capacity of over 80%: Asn, His, Phe, Trp and Tyr. In this case, Cys, compared to all AAs, has a significant activity and not the highest, contrary to what would have been expected (**Figure 3.1**).

The potential for inhibition of NO radicals by AAs is over 51%, highlighting the same AA as in the case of the SORS method, joining Cys. The highest activity was obtained for Tyr around 58%.

3.3.2. Reducing antioxidant power

According to the FRAP analysis, Cys (96.72%) and Trp (85.30%) have the highest free radical reduction power of all the AAs investigated.

However, after testing the potential of reducing power of free radicals, the best results were observed for Cys (95.77%) for CUPRAC assay and for Arg, Gln, Leu and Trp for CHROMAC assay, as can be seen in **Figure 3.1**.

Of all the AAs studied, only three of them showed significant antioxidant activity by all methods applied: Cys, Trp and Tyr.

5.3.3. Chelating capacity

The chelating capacity of metal ions, especially ferrous ions, of proteinogenic AA is moderate, except for Arg (92.15%), Gly (65.27%), Glu (63.89%), Tyr (66.40 %) and Val (68.74%)) as can be seen in **Figure 3.1**. However, Arg and Asn have the highest chelating capacity compared to the other AAs investigated.

In conclusion, Arg, Asn and Tyr can be considered chelators of metal ions and reducers of oxidative stress by supplementation.

3.3.4. Cluster analysis and pricipal component analysis

The dendrogram (**Figure 3.1**) was obtained by applying the HCA analysis on the matrix of data corresponding to the antioxidant and chelating capacity of the investigated AA (using the covariance matrix), two well-defined groups of methods can be highlighted. The first group includes the DPPH, NO, CHROMAC and SORS methods, including chelating capacity, and the second group contains the ABTS, CUPRAC and FRAP methods.

These results are well supported by the graphical representation Heat Map where both (di)similarities between AA and (di)similarities between applied methods can be observed simultaneously (**Figure 3.2**). The highest values appear for DPPH and SORS and the lowest for the ABTS and CUPRAC methods.

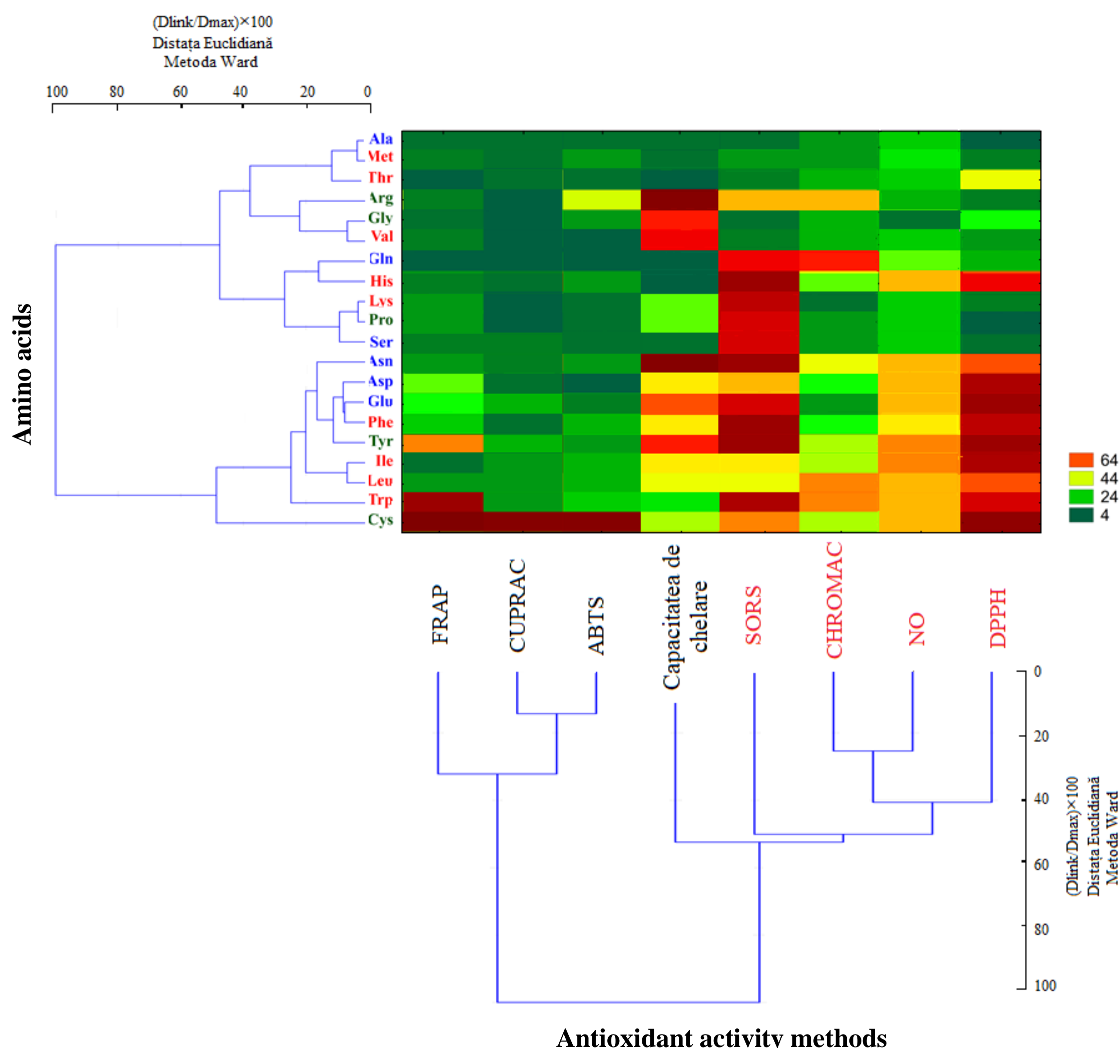


Figure 3.1. Combined dendrogram [(Dlink/Dmax)*100 –Ward's method] corresponding to all antioxidant and chelating capacity method sand dendrogram [(Dlink/Dmax)*100 –Ward's method] corresponding to the investigated amino acids with the heat map corresponding to amino acids and all method.

Figure 3.1 shows the similarities and differences between AAs according to all methods applied in this study. Thus, it can be seen that AAs form two well-defined groups: the group of AAs with significant activity and the group of AAs with low activity. Cysteine being the only AA that differs from the rest by its antioxidant activity.

By graphically representation of scores using the first three principal components (**Figure 3.2**), a satisfactory separation of AAs is obtained depending on the antioxidant potential, where it can be seen that Cys appears as an extreme.

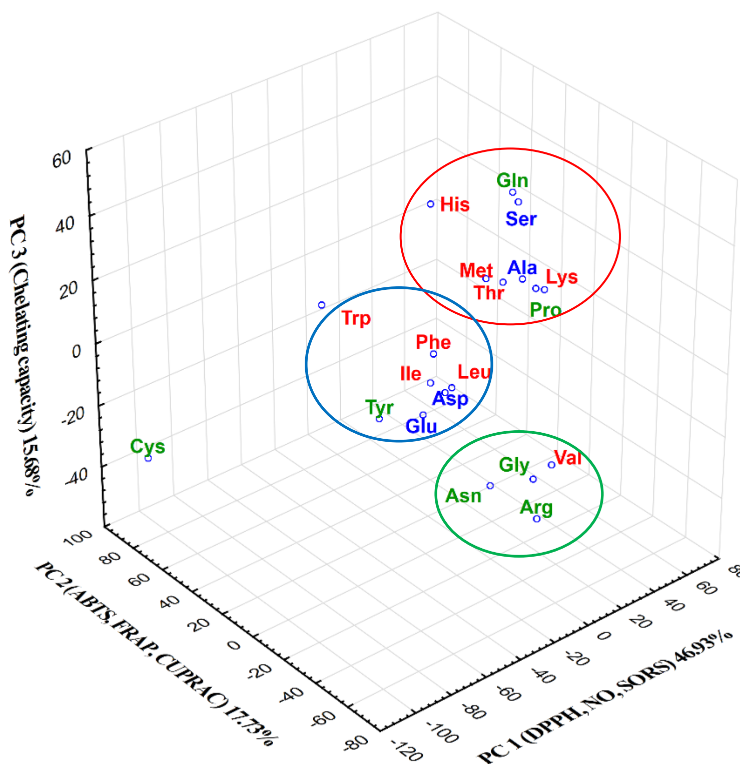


Figure 3.2. 3D-score scatterplot, corresponding to the investigated amino acids (PC1—DPPH, NO, SORS—46.93%, PC2—ABTS, FRAP, CUPRAC—17.73%, PC3—chelating capacity—15.68%), obtained after applying PCA on the covariancematrix of the all dataset.

The vector corresponding to the first principal component (PC1) can be associated with the radical scavenging capacity—DPPH, NO, SORS, the vector of the second principal component (PC2)—the reduction of antioxidant power—ABTS, FRAP, CUPRAC and the vector of the third principal component (PC3)—chelation capacity of metal ions. A very good separation can be observed in 3 well-defined groups of AAs along to PC3——the component associated with chelating capacity of metal ions.

5.3.5. SRD analysis

The results shown that the method closest to the “average value” is the DPPH method, while the most different method is the ABTS method (**Figure 3.3**) As can be seen in the figure, the methods that appear the left of the Gaussian curve are most similar to the "average value". Thus, the performance of the applied methods can be visualized (from left to right).

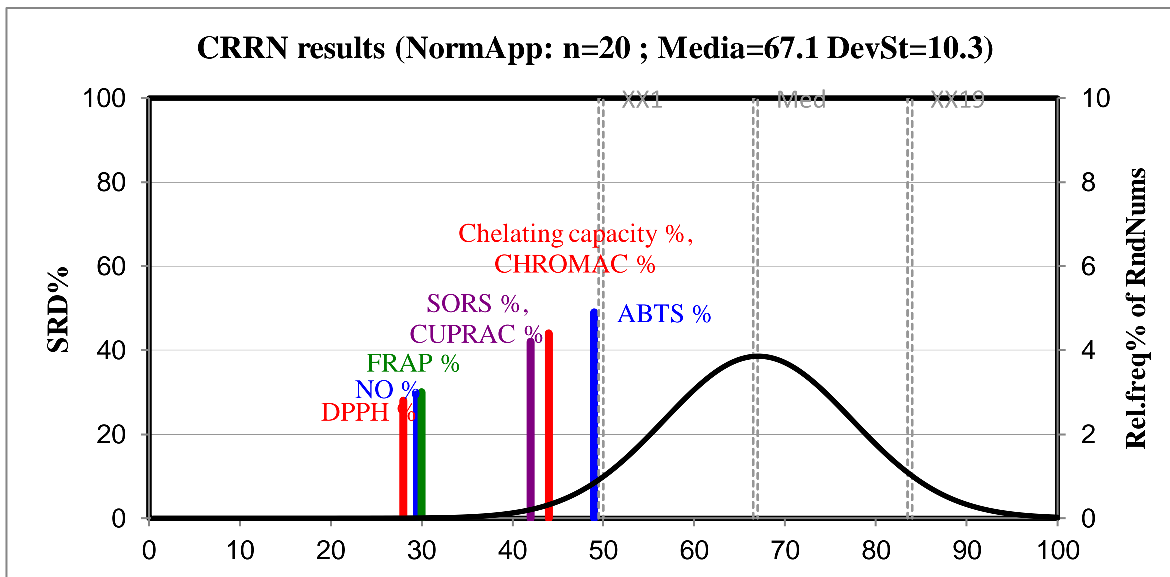


Figure 3.3. SRD-CRRN test results of the data matrix. Average was used as a golden standard. Scaled SRD values are plotted on the x-axis and left y-axis, the right y-axis shows the relative frequencies (black curve). Parameters of the Gaussian fit: media = 67, standard deviation = 10.3 Probability levels 5% (XX1), Median (Med), and 95% (XX19) are also given.

4. Determination of antioxidant activity and chelating capacity of some biogenic amines and related drugs

4.1. Introducion

The purpose of this chapter is to evaluate the antioxidant potential of catecholamines and drugs with similar structure, such as scavenging radical capacity (DPPH, ABTS, SORS and NO), reduction of antioxidant power (FRAP, CUPRAC and CHROMAC) and chelation of metals ions. And at the same time, interpreting and comparing the results obtained by applying advanced chemometric methods such as HCA, PCA and SRD.

4.2. Materials and methods

4.2.1 Samples

The biogenic amines investigated in this study include adrenaline, noradrenaline, dopamine, and related drugs: methyldopa, L-dopa, D-dopa, metaraminol, ritodrine, adrenalone, albuterol, metaproterenol, terbutaline, isoprenaline, and methoxamine of analytical grade obtained from Merck and Sigma (Sigma-Aldrich GmbH, Sternheim, Germany). All other reagents were either of analytical grade or of the highest quality available.

The stock solutions of all compounds, including positive controls (quercetin and trolox) were prepared in absolute ethanol (96%) at concentration 1 mM. EDTA solution was prepared in water at concentration 1mM. For those related drugs that were poorly soluble in ethanol (adrenaline, D-dopa), stoichiometric hydrochloric acid was added. The final concentration of investigated compounds (300 μL) was: 1.67 μM for DPPH, ABTS, SORS, FRAP, CUPRAC, FIC assays; 20 μM for NO assay and 16.67 μM for CHROMAC assay. The same concentrations are for reference antioxidant solutions.

4.2.2. Methods of determination of antioxidant activity

For derermination of antioxidant activity were used eight methods groups as radical scavenging assays (DPPH, ABTS, SORS și NO), reducing antioxidant power assays (FRAP, CUPRAC și CHROMAC) and metal chelating capacity.

4.2.3. Chemometric methods

For realization of purpose, were used the following chemometric methods for multidimensional data analysis, such as HCA, PCA and SRD analysis.

4.3. Results and discussion

The results obtained after the application of 8 methods for determining the antioxidant activity combined with chemometric methods are summarized graphically in **Figure 4.1.**

Regarding the results obtained by the methods of quenching the power of free radicals, most of the compounds showed a significant activity except for some drugs (ritodrine, methoxamine, D,L-dopa and adrenaline). The results regarding the reduction of antioxidant power, several groups with very close values can be identified: norepinephrine and isoprenaline; dopamine, adrenaline, methyldopa; albuterol, metaproterenol, terbutaline; L-dopa and D-dopa; metaraminol and metoxamine having the lowest values in all three methods. Adrenaline and ritodrine appear quite different again. The highest reducing power, according to the FRAP method, has: adrenaline (96.02%) methyldopa (95.97%), dopamine (94.67%) and isoprenaline (93.72%), and the lowest—metaraminol (5.79%).

The chelating capacity of Fe^{2+} ions of the analyzed compounds is low but significant except for adrenalone (66.35%), metaraminol (55.31%), metaproterenol (49.58%) and terbutaline (45.64%), which showed a capacity chelation slightly above 50%. The lowest value, in this case, was obtained for methyldopa (0.28%).

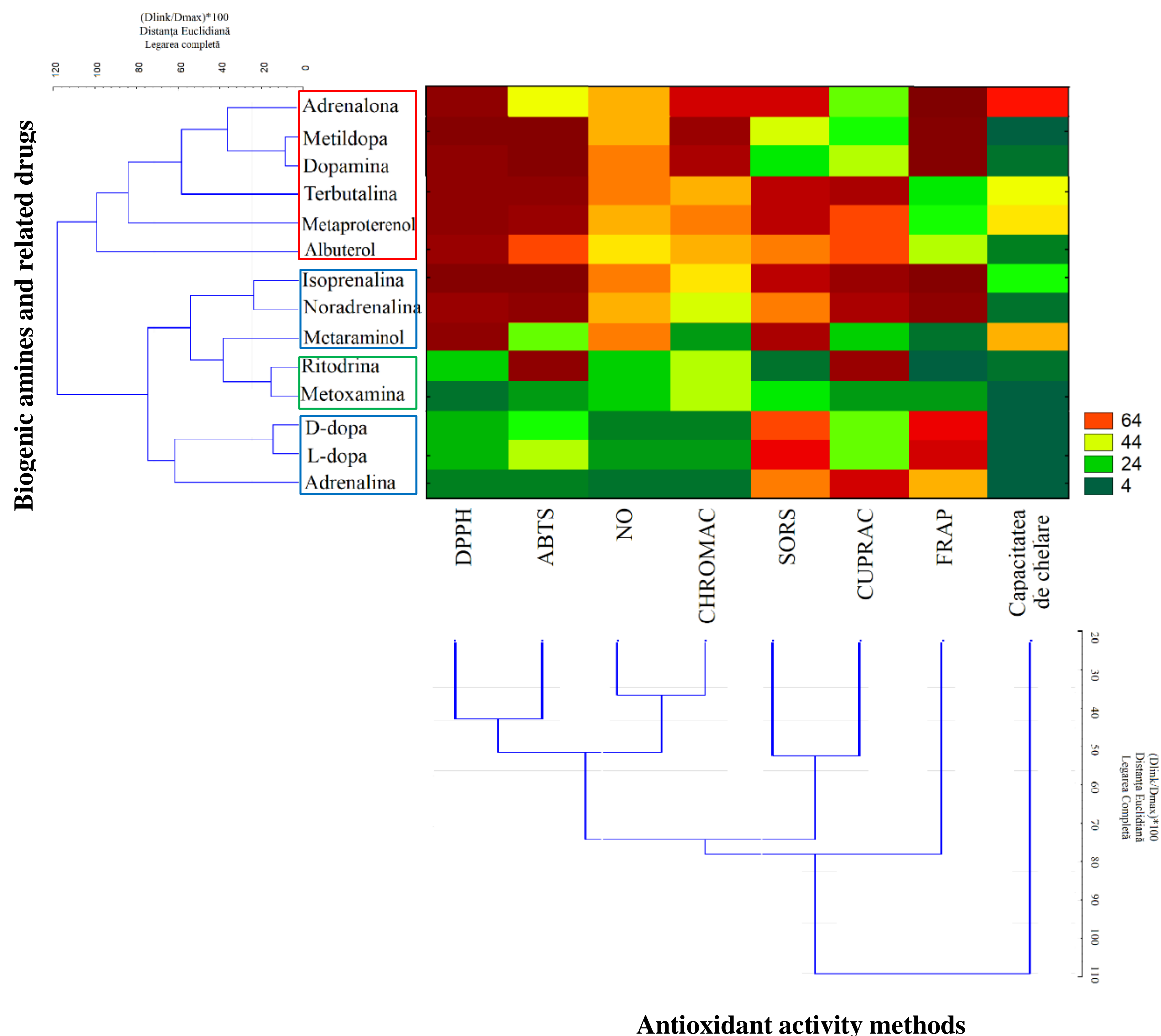


Figure 4.1. The heat map corresponding to all investigated compounds and all assays coupled with hierarchical cluster analysis.

By careful examination of **Figure 4.1** (based on color intensity) it is very interesting to note that most adrenergic drugs have the highest antioxidant activity, except for Metoxamine, L-dopa and D-dopa (which has no pharmacological activity). In the case of catecholamines, the results obtained for noradrenaline are much more similar to dopamine and adrenergic drugs, but quite different from adrenaline.

The results obtained by applying PCA on the unmodified data matrix (14 samples × 8 methods) indicate a significant reduction of variables. The first three principal components take over 84.33% of the total variation, the first two components take over only 68.18% of the total variance, and the first—48.80%. However, the 3D representation of the scores corresponding to the first three main components indicates a satisfactory separation of the compounds according to their antioxidant activity.

By representing the first three components, the vector corresponding to the first component (PC1) can be associated with the radical scavenging power—DPPH, ABTS and NO, the vector of the second principal component (PC2)—the radical scavenging power—ABTS, FRAP and CUPRAC and the vector of the third component (PC 3)—chelation capacity of metal ions.

According to SRD, the best method to express antioxidant activity of investigated biogenic amines and their related drugs is DPPH because it is the method closest to the "average antioxidant and chelating capacity'', while the FRAP method is the most different. The results are well illustrated in **Figure 4.2**. We have also to remark the high similarity of the clusters with the dendrogram obtained applying HCA.

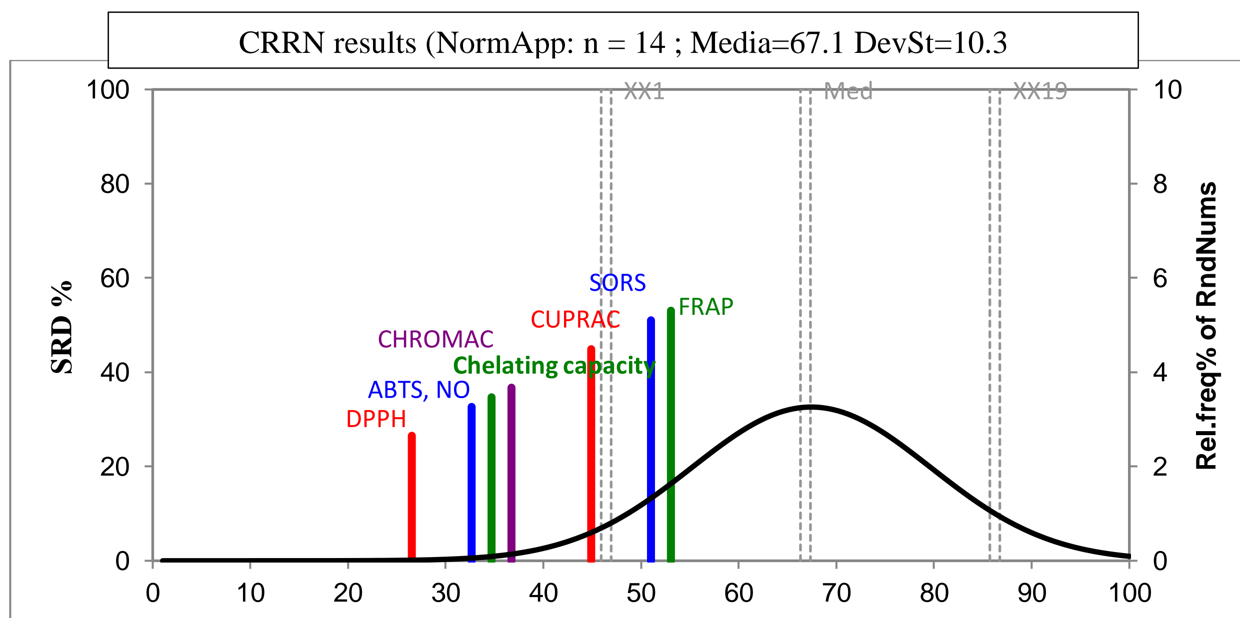


Figure 4.2. Evaluation of the eight assays using the sum of ranking differences. Average was used as a golden standard.

5. Modeling of amino acids lipophilicity using QSPR studies

5.1. Introduction

The objective of this chapter was to identify a mathematical models for correct estimation and prediction of lipophilicity of AAs, starting from the complex information provided by their molecular structure. Currently the literature abounds with experimental data on the lipophilicity of AAs determined in the flask and on various chromatographic plates, but the major objectives of this study are to create a model "as general as possible" that can be applied to determine lipophilicity for a large classes of compounds with similar structure of AAs, as well as the identification of the best method of expressing lipophilicity.

Prediction models were generated by multiple linear regression (MLR) using genetic algorithms for the selection of initial variables. The best prediction models were selected based on a wide set of classic quality statistical parameters as well as a series of newly developed and applied fuzzy quality parameters, which differ in the way of calculating the distance from the regression line (thus minimizing the contribution of residues or minimizing the difference between $\bar{y}_i$ estimated and y_i measured). A comparison between classic and fuzzy models was presented.

5.2. Materials and methods

5.2.1. Structure of amino acids

The AA's structures were represented using Chem3D Ultra 8.0 and optimized using the MM+ (Molecular Mechanics Force Field) procedure included in the same program. The optimized geometric structures were used in the Dragon plus 5.4 program and the SMILES structures in the Alchemy program (http://www.tripos.com) in order to calculate the molecular descriptors.

5.2.2. Methods

In this study, the development of calculated and predicted models was performed using MobiDigs v.1.0 [30] using multiple linear regression method (MLR). The selection of the most significant

variables (descriptors) was performed using genetic algorithms [31] methodology, included in the same program.

The development of fuzzy calculated and predicted models was done using the *Sadic* program.

5.3. Results and discussion

Multiple linear regression models for estimating lipophilicity parameters of proteinogenic AAs were obtained by selecting the most significant descriptors using the genetic algorithm methodology (RLM-GA) using the chromatographic lipophilicity coefficient (R_{M0}) obtained in a previous experimental study [32] for 16 proteinogenic amino acids (Ala, Arg, Asn, Asp, Cys, Glu, Gly, His, Leo, Lys, Met, Phe, Pro, Ser, Tyr, Val). Because, at that time, the lipophilicity coefficient R_{M0} was not determined experimentally for 4 amino acids (Gln, Ile, Thr, Trp), they were used as a test set and external validation of the created models.

Starting with the independent variables obtained by calculating the theoretical descriptors using DRAGON and Alchemy software, the most statistically significant RLM models with 3, 4 and 5 independent variables were generated using genetic algorithms methodology. The best models were selected according to the values of the statistical quality parameters. The descriptors, obtained with both software, retained in the models are presented in **Table 5.1**. The best models obtained using the descriptors with DRAGON and Alchemy with the highest predictive power together with the statistical quality parameters are presented in **Table 5.2**.

The statistical parameters that correspond to all regression models illustrate a statistically significant to moderate prediction power. At the same time, it can be seen that the most powerful models are those containing 5 or 4 molecular descriptors.

Analyzing the data from **Table 5.2** it can be seen that an R^2 of over 89% was obtained for Alchemy models and for models containing Dragon descriptors of over 99%.

On closer examination, it can be seen that the descriptors that most strongly correlate with the chromatographic lipophilicity coefficient (R_{M0}) in the case of Alchemy set are: molecular (Polar) and specific (Pol.Pol) polarization, the sum of the absolute values on nitrogen and oxygen atoms in the molecule ($ABSQ_{ON}$), volume and Wiener index (WienI).

Table 5.1. Retained descriptors in RLM-GA models and calculated with Alchemy[2000] and Dragon Plus 5.4 software.

AA	Data*	Descriptors																
		ALCHEMY							DRAGON									
	R_{M0}	**Volume**	**ABSQon**	**MaxQ⁻**	**Polar**	**Sp.Pol**	$^{0}\chi^{v}$	**WienI**	**MATS3p**	**MATS3v**	**DP11**	**DP12**	**RDF070e**	**Mor12u**	**Ap**	**HATS2u**	**SP13**	**H1p**
Ala	−1.14	164.62	1.92	−0.35	17.49	0.11	6.73	247.00	−0.27	−0.34	0.02	0.01	0.00	−0.55	2.66	0.61	0.002	0.40
Arg	−0.60	165.16	1.92	−0.36	17.87	0.11	6.71	247.00	−0.06	−0.06	6.73	6.42	5.70	−1.43	11.46	0.32	6.06	0.51
Asn	−1.19	114.90	1.66	−0.42	11.88	0.10	4.70	96.00	−0.31	−0.32	1.50	0.95	0.25	−1.01	4.76	0.43	0.66	0.43
Asp	−1.26	112.39	1.57	−0.33	11.04	0.10	4.57	96.00	−0.36	−0.38	1.42	0.89	0.00	−0.82	4.22	0.42	0.60	0.42
Cys	−0.90	101.85	0.94	−0.33	11.48	0.11	4.56	46.00	0.01	−0.04	0.43	0.19	0.00	−0.52	3.69	0.55	0.10	0.56
Gln	–	132.91	1.66	−0.42	13.71	0.10	5.41	136.00	−0.21	−0.23	3.30	2.73	1.42	−1.03	5.74	0.39	2.28	0.47
Glu	−1.18	129.38	1.57	−0.33	12.87	0.10	5.28	136.00	−0.16	−0.18	3.38	2.81	1.92	−1.20	6.39	0.39	2.37	0.46
Gly	−1.07	68.43	0.95	−0.33	6.64	0.10	2.64	18.00	0.34	0.23	0.01	0.004	0.00	−0.34	1.45	0.77	0.001	0.24
His	−0.59	137.90	1.53	−0.33	15.43	0.11	5.82	165.00	−0.01	−0.01	2.74	2.11	3.22	−1.06	7.31	0.42	1.65	0.68
Ile	–	135.50	0.94	−0.33	13.86	0.10	5.80	92.00	0.14	0.11	1.50	0.98	2.46	−0.77	6.61	0.41	0.72	0.55
Leu	−0.47	135.82	0.94	−0.33	13.86	0.10	5.79	96.00	0.19	0.15	0.90	0.47	1.61	−1.09	6.46	0.45	0.24	0.65
Lys	−0.93	148.91	1.27	−0.33	15.21	0.10	5.92	143.00	0.08	0.06	5.10	4.69	3.21	−1.15	8.26	0.37	4.29	0.48
Met	−0.55	137.48	0.94	−0.33	15.02	0.11	6.15	102.00	0.24	0.19	3.79	3.27	3.52	−0.52	7.25	0.51	2.83	0.56
Phe	−0.02	157.90	0.94	−0.33	18.14	0.11	6.60	212.00	0.33	0.30	3.96	3.40	4.33	−1.04	10.37	0.36	2.94	0.86
Pro	−0.90	108.70	0.93	−0.33	11.24	0.10	4.55	62.00	0.21	0.15	0.28	0.11	0.00	−0.97	4.06	0.55	0.05	0.53
Ser	−1.21	93.57	1.33	−0.39	9.12	0.10	3.66	46.00	−0.40	−0.46	0.23	0.09	0.00	−0.68	2.98	0.54	0.04	0.39
Thr	–	110.39	1.33	−0.39	10.83	0.10	4.54	65.00	−0.15	−0.19	0.27	0.11	0.00	−0.56	4.14	0.45	0.06	0.48
Trp	–	186.60	1.22	−0.33	22.32	0.12	8.10	396.00	0.19	0.18	5.87	5.48	6.88	−1.01	16.43	0.33	5.26	0.89
Tyr	−0.44	165.64	1.34	−0.39	18.78	0.11	6.97	268.00	0.11	0.08	5.39	4.99	4.14	−1.20	10.69	0.33	4.63	0.73
Val	−0.68	118.81	0.94	−0.33	12.02	0.10	5.09	65.00	0.01	−0.01	0.30	0.12	0.00	−0.74	5.15	0.46	0.07	0.51

* R_{M0} experimental datas obtained on HPTLC RP-18W chromatographic plates extracted from [32]

Table 5.2. Multiple linear regression models obtained for the prediction of the retention coefficient (R_{M0}) by applying the methodology of genetic algorithms to Alchemy and Dragon descriptors.

ID	Dimension	Models	R^2 %	F	*s*	RSS	SDEC	Q^2 %	PRESS	SDEP
ALCHEMY										
A	5	$^{*}R_{M0} = 14.01 - 0.14\cdot\mathbf{Volume} + 1.68\cdot\mathbf{MaxQ^{-}} + 1.54\cdot\mathbf{Polar} - 155.52\cdot\mathbf{Sp.Pol} - 0.008\cdot\mathbf{Wienl}$ $^{**}R_{M0} = -11.08\cdot\mathbf{Volume} + 0.15\cdot\mathbf{MaxQ^{-}} + 15.01\cdot\mathbf{Polar} - 2.61\cdot\mathbf{Sp.Pol} - 1.691\cdot\mathbf{WienI}$	**89.19**	16.51	**0.143**	**0.206**	**0.113**	**76.37**	**0.450**	**0.168**
B	4	$^{*}R_{M0} = 13.38 - 0.14\cdot\mathbf{Volume} + 1.56\cdot\mathbf{Polar} - 156.11\cdot\mathbf{Sp.Pol} - 0.008\cdot\mathbf{WienI}$ $^{**}R_{M0} = -11.06\cdot\mathbf{Volume} + 15.14\cdot\mathbf{Polar} - 2.62\cdot\mathbf{Sp.Pol} - 1.86\cdot\mathbf{WienI}$	**87.33**	18.95	0.148	0.242	0.123	**73.03**	0.514	0.179
C	3	$^{*}R_{M0} = -0.97 - 0.68\cdot\mathbf{ABSQ_{ON}} + 0.18\cdot\mathbf{Polar} - 0.28\cdot{}^{\mathbf{0}}\chi^{\mathbf{v}}$ $^{**}R_{M0} = -0.70\cdot\mathbf{ABSQ_{ON}} + 1.79\cdot\mathbf{Polar} - 0.94\cdot{}^{\mathbf{0}}\chi^{\mathbf{v}}$	83.45	20.16	0.162	0.315	0.140	71.50	0.543	0.184
D	2	$^{*}R_{M0} = -1.17 - 0.66\cdot\mathbf{ABSQ_{ON}} + 0.09\cdot\mathbf{Polar}$ $^{**}R_{M0} = -0.68\cdot\mathbf{ABSQ_{ON}} + 0.86\cdot\mathbf{Polar}$	80.95	27.62	0.167	0.363	0.151	68.96	0.591	0.192
DRAGON										
E	5	$^{*}R_{M0} = -1.36 + 0.36\cdot\mathbf{MATS3p} - 0.22\cdot\mathbf{DP11} + 0.09\cdot\mathbf{RDF070e} + 0.37\cdot\mathbf{Mor12u} + 0.20\cdot\mathbf{Ap}$ $^{**}R_{M0} = 0.24\cdot\mathbf{MATS3p} - 1.37\cdot\mathbf{DP11} + 0.50\cdot\mathbf{RDF070e} + 0.32\cdot\mathbf{Mor12u} + 1.69\cdot\mathbf{Ap}$	**99.83**	**1172.90**	**0.018**	**0.003**	**0.014**	**99.54**	**0.009**	**0.023**
F	4	$^{*}R_{M0} = -3.05 - 0.22\cdot\mathbf{DP12} + 0.31\cdot\mathbf{Mor12u} + 0.32\cdot\mathbf{Ap} + 2.07\cdot\mathbf{HATS2u}$ $^{**}R_{M0} = -1.33\cdot\mathbf{DP12} + 0.27\cdot\mathbf{Mor12u} + 2.73\cdot\mathbf{Ap} + 0.69\cdot\mathbf{HATS2u}$	**99.06**	290.00	0.040	0.018	0.034	**98.17**	0.035	0.047
G	3	$^{*}R_{M0} = -3.32 - 0.24\cdot\mathbf{SP13} + 0.32\cdot\mathbf{Ap} + 2.85\cdot\mathbf{HATS2u}$ $^{**}R_{M0} = -1.30\cdot\mathbf{SP13} + 2.71\cdot\mathbf{Ap} + 0.95\cdot\mathbf{HATS2u}$	**97.51**	156.50	0.063	0.047	0.055	**95.56**	0.085	0.073
H	2	$^{*}R_{M0} = -1.64 + 0.61\cdot\mathbf{MATS3v} + 1.61\cdot\mathbf{H1p}$ $^{**}R_{M0} = 0.40\cdot\mathbf{MATS3v} + 0.68\cdot\mathbf{H1p}$	88.45	49.80	0.130	0.220	0.117	84.97	0.286	0.134

*—standardized coefficients

**—non-standardized coefficients

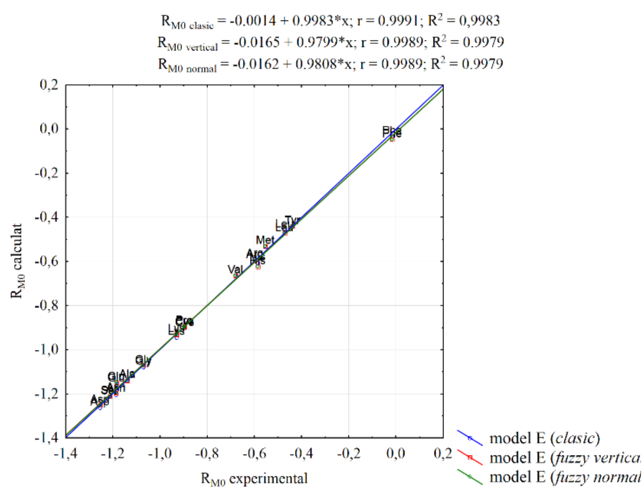
Table 5.3. Regression coefficients and fuzzy quality parameters obtained for prediction of R_{M0} on Dragon and Alchemy descriptors applying RLM-GA.

ID	Dimension	Type	Models	QC_1	QC_3	QC_5	**QC_6**	NQC_5	**NQC_6**
ALCHEMY									
A	5	**Clasic**	R_{M0} = 14.01 – 0.14·**Volume** + 1.68·**MaxQ⁻** + 1.54·**Polar** – 155.52·**Sp.Pol** – 0.008·**WienI**	0.643	0.212	1.747	5.216	0.437	0.326
		Vertical	R_{M0} = 14.45 – 0.14·**Volume** + 1.84·**MaxQ⁻** + 1.59·**Polar** – 163.10·**Sp.Pol** – 0.01·**WienI**	1.038	0.245	1.691	**6.466**	0.423	**0.404**
		Normal	R_{M0} = 20.02 – 0.19·**Volume** + 1.08·**MaxQ⁻** + 2.14·**Polar** – 223.29·**Sp.Pol** – 0.01·**WienI**	1.883	0.203	1.275	6.314	0.319	0.395
B	4	**Clasic**	R_{M0} =13.38 – 0.14·**Volume** + 1.56·**Polar** – 156.11·**Sp.Pol** – 0.008·**WienI**	0.690	0.224	1.828	4.716	0.457	0.295
		Vertical	R_{M0} = 17.62 – 0.18·**Volume** + 1.96·**Polar** – 201.66·**Sp.Pol** – 0.012·**WienI**	1.178	0.269	1.839	**5.909**	0.460	**0.369**
		Normal	R_{M0} = 25.10 – 0.22·**Volume** + 2.47·**Polar** – 280.90·**Sp.Pol** – 0.014·**WienI**	1.070	0.231	1.671	5.793	0.418	0.362
C	3	**Clasic**	R_{M0} = −0.97 – 0.68·**ABSQ$_{ON}$** + 0.18·**Polar** – 0.28·$^{0}\chi^{V}$	0.344	0.229	2.151	4.912	0.538	0.307
		Vertical	R_{M0} = −0.89 – 0.70·**ABSQ$_{on}$** + 0.23·**Polar** – 0.40·$^{0}\chi^{V}$	0.356	0.233	2.563	4.385	0.641	0.274
		Normal	R_{M0} = −0.87 – 0.71·**ABSQ$_{on}$** + 0.23·**Polar** – 0.41·$^{0}\chi^{V}$	0.233	0.183	2.090	**5.160**	0.523	**0.322**
D	2	**Clasic**	R_{M0} = −1.17 – 0.66·**ABSQ$_{ON}$** + 0.09·**Polar**	0.339	0.24	2.093	4.851	0.523	0.303
		Vertical	R_{M0} = −0.80 – 0.83·**ABSQ$_{ON}$** + 0.08·**Polar**	0.353	0.241	2.297	4.679	0.574	0.292
		Normal	R_{M0} = −1.64 + 0.76·**ABSQ$_{ON}$** + 1.61·**Polar**	0.324	0.225	1.557	**5.588**	0.389	**0.349**
DRAGON									
E	5	**Clasic**	R_{M0} = −1.36 + 0.36·**MATS3p** – 0.22·**DP11** + 0.09·**RDF070e** + 0.37·**Mor12u** + 0.20·**Ap**	0.163	0.03	2.152	4.73	0.538	0.296
		Vertical	R_{M0} = −1.36 + 0.37·**MATS3p** − 0.21 * **DP11** + 0.08·**RDF070e** + 0.38·**Mor12u** + 0.20·**Ap**	0.174	0.026	1.871	4.975	0.468	0.311
		Normal	R_{M0} = −1.36 + 0.37−**MATS3p** − 0.21 * **DP11** + 0.08·**RDF070e** + 0.38·**Mor12u** + 0.20·**Ap**	0.168	0.021	1.785	**6.433**	0.446	**0.402**
F	4	**Clasic**	R_{M0} = −3.05 – 0.22·**DP12** + 0.31·**Mor12u** + 0.32·**Ap** + 2.07·**HATS2u**	0.185	0.059	1.619	4.846	0.405	0.303
		Vertical	R_{M0} = −3.07 − 0.23·**DP12** + 0.32·**Mor12u** + 0.33·**Ap** + 2.10·**HATS2u**	0.191	0.06	1.852	4.98	0.463	0.311
		Normal	R_{M0} = −3.24 − 0.25·**DP12** + 0.30·**Mor12u** + 0.34·**Ap** + 2.30·**HATS2u**	0.082	0.048	1.282	**5.959**	0.321	**0.372**
G	3	**Clasic**	R_{M0} = −3.32 – 0.24·**SP13** + 0.32·**Ap** + 2.85·**HATS2u**	0.229	0.092	2.599	4.451	0.65	0.278
		Vertical	R_{M0} = −3.40 − 0.19·**SP13** + 0.29·**Ap** + 2.47·**HATS2u**	0.257	0.109	2.311	4.608	0.578	0.288
		Normal	R_{M0} = −4.37 − 0.26·**SP13** + 0.36·**Ap** + 3.82·**HATS2u**	0.144	0.102	1.873	**5.735**	0.468	**0.358**
H	2	**Clasic**	R_{M0} = −1.64 + 0.61·**MATS3v** + 1.61·**H1p**	0.274	0.167	1.905	4.695	0.476	0.293
		Vertical	R_{M0} = −1.63 + 0.76·**MATS3v** + 1.59·**H1p**	0.25	0.163	1.888	4.940	0.472	0.309
		Normal	R_{M0} = −1.65 + 0.76·**MATS3v** + 1.616·**H1p**	0.184	0.153	1.782	**5.435**	0.446	**0.340**

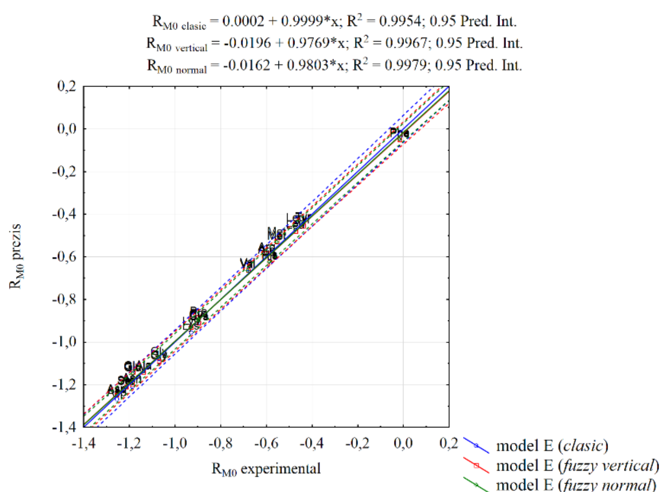
The most significant descriptors calculated with Dragon software are radial distribution descriptors (RDF), autocorrelation indices and geometric descriptors related mainly to atomic polarizability (MATS3p, Ap, H1p), van der Waals atomic volume (MATS3v), Sanderson atomic electronegativity (RDF070e) and molecular Randice profiles (DP11, DP12) .

To examine the ability of the obtained models to predict the chromatographic lipophilicity coefficient, the calculated and predicted R_{M0} values in the validation process were compared with the experimental ones. Good correlation values were found for training dataset (16 amino acids) and the training and test set (20 amino acids) for models using Dragon descriptors (**Figure 5.1**) and Alchemy descriptors (**Figure 5.2**). Based on the classical quality coefficients, the best model that can predict lipophilicity with a power of 99.83% and which explains 99.83% of the total variance, was obtained for models containing Dragon descriptors and a power of 76.37% for models containing descriptors Alchemy.

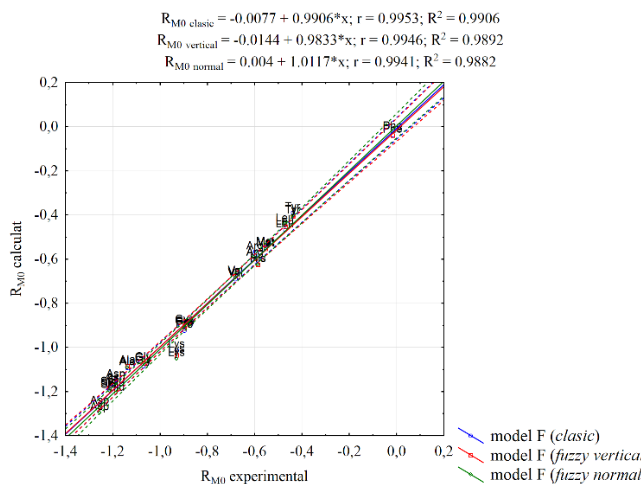
A comparison between classical and fuzzy models can be seen in **Figure 5.3** and **Table 5.4** where all classical quality coefficients for each fuzzy model are summarized. Also, it can be seen that fuzzy models show a significant improvement in terms of predictive power, for both models (that contain Alchemy and Dragon descriptors.

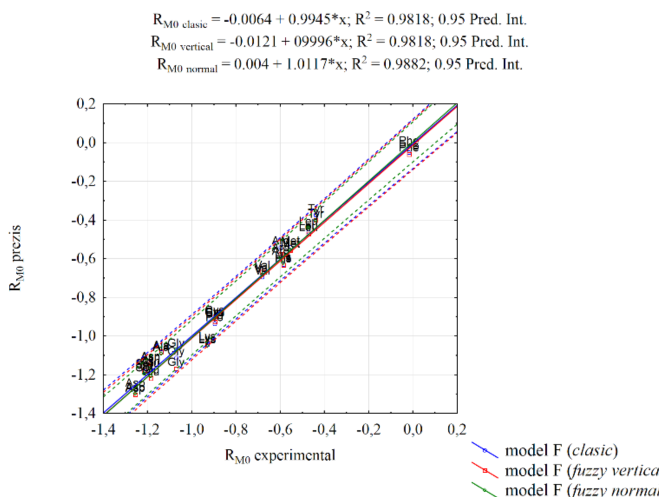


(a)



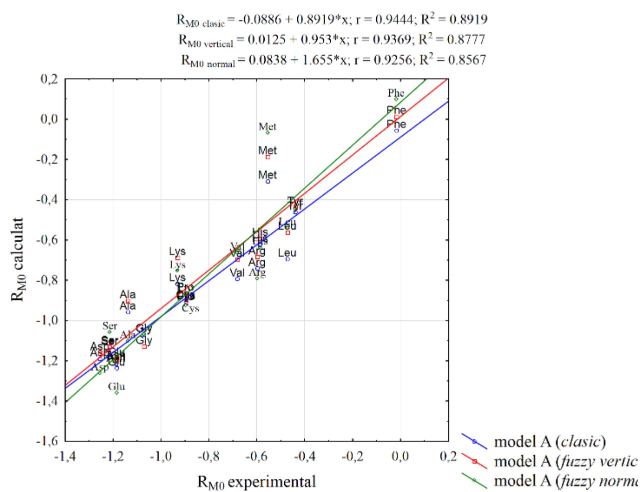
(b)

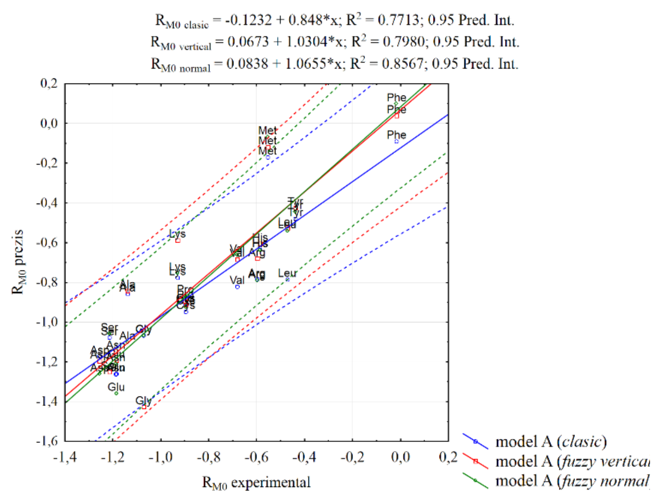




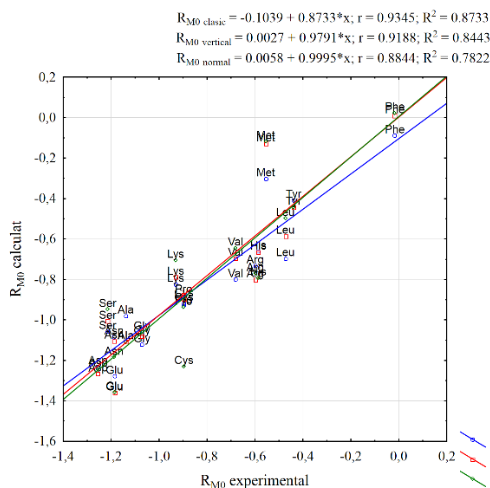
(c) (d)

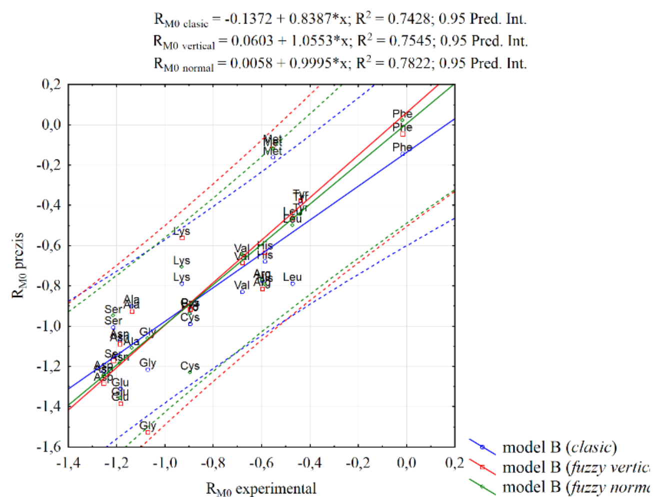
Figure 5.1. Correlations between experimental and calculated (a, c) and experimental andpredicted (b, d) R_{M0} values for models E (a, b) and F (c, d) that contain Dragon descriptors.





(a) (b)





(c) (d)

Figure 5.2. Correlations between experimental and calculated (a, c), experimental and predicted (b, d) R_{M0} values for models A (a, b) and B (c, d) that contain Alchemy descriptors.

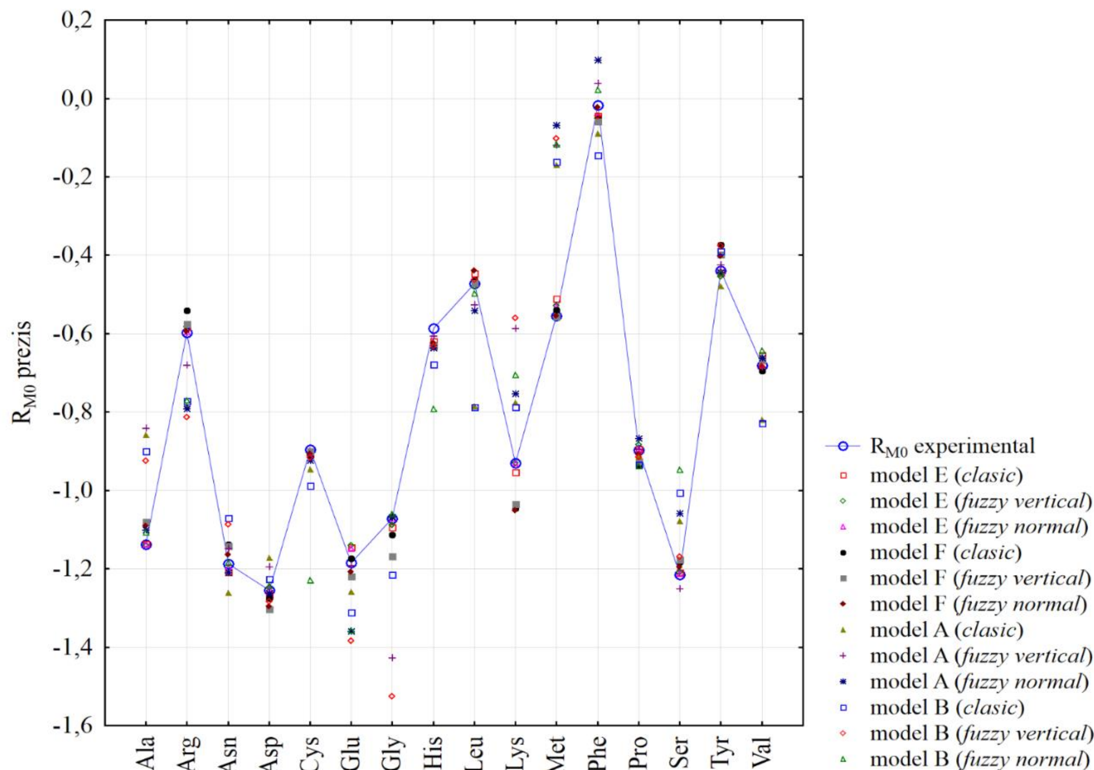


Figura 5.3. Predicted R_{M0} values for the best models in the external cross validation process.

Table 5.4. Comparison of classic RLM-GA models and new fuzzy RLM-GA models developed using Dragon and Alchemy descriptors.

ID	Dimension	Type	Models	R^2 %	F	*s*	RSS	SDEC	Q^2 %	PRESS	SDEP
			ALCHEMY								
A	5	**Clasic**	$R_{M0} = 14.01 - 0.14\cdot \mathbf{Volume} + 1.68\cdot \mathbf{MaxQ^{-}} + 1.54\cdot \mathbf{Polar} - 155.52\cdot \mathbf{Sp.Pol} - 0.008\cdot \mathbf{WienI}$	89.19	16.51	0.143	0.206	0.113	76.37	0.450	0.168
		Vertical	$R_{M0} = 14.45 - 0.14\cdot \mathbf{Volume} + 1.84\cdot \mathbf{MaxQ^{-}} + 1.59\cdot \mathbf{Polar} - 163.10\cdot \mathbf{Sp.Pol} - 0.01\cdot \mathbf{WienI}$	84.94	14.028	0.169	0.287	0.134	71.53	0.543	0.184
		Normal	$R_{M0} = 20.02 - 0.19\cdot \mathbf{Volume} + 1.08\cdot \mathbf{MaxQ^{-}} + 2.14\cdot \mathbf{Polar} - 223.29\cdot \mathbf{Sp.Pol} - 0.01\cdot \mathbf{WienI}$	79.83	13.212	0.196	0.384	0.155	**79.83**	**0.384**	**0.155**
B	4	**Clasic**	$R_{M0} = 13.38 - 0.14\cdot \mathbf{Volume} + 1.56\cdot \mathbf{Polar} - 156.11\cdot \mathbf{Sp.Pol} - 0.008\cdot \mathbf{WienI}$	87.33	18.95	0.148	0.242	0.123	73.03	0.514	0.179
		Vertical	$R_{M0} = 17.62 - 0.18\cdot \mathbf{Volume} + 1.96\cdot \mathbf{Polar} - 201.66\cdot \mathbf{Sp.Pol} - 0.012\cdot \mathbf{WienI}$	81.94	17.69	0.177	0.344	0.147	63.28	0.699	0.209
		Normal	$R_{M0} = 25.10 - 0.22\cdot \mathbf{Volume} + 2.47\cdot \mathbf{Polar} - 280.90\cdot \mathbf{Sp.Pol} - 0.014\cdot \mathbf{WienI}$	72.14	12.612	0.220	0.531	0.182	72.14	0.531	0.182
C	3	**Clasic**	$R_{M0} = -0.97 - 0.68\cdot \mathbf{ABSQ_{ON}} + 0.18\cdot \mathbf{Polar} - 0.28\cdot {}^{0}\chi^{V}$	83.45	20.16	0.162	0.315	0.140	71.50	0.543	0.184
		Vertical	$R_{M0} = -0.89 - 0.70\cdot \mathbf{ABSQ_{on}} + 0.23\cdot \mathbf{Polar} - 0.40\cdot {}^{0}\chi^{V}$	82.79	20.776	0.165	0.328	0.144	71.72	0.540	0.184
		Normal	$R_{M0} = -0.87 - 0.71\cdot \mathbf{ABSQ_{on}} + 0.23\cdot \mathbf{Polar} - 0.41\cdot {}^{0}\chi^{V}$	82.77	20.981	0.165	0.328	0.143	**82.77**	**0.328**	**0.143**
D	2	**Clasic**	$R_{M0} = -1.17 - 0.66\cdot \mathbf{ABSQ_{ON}} + 0.09\cdot \mathbf{Polar}$	80.95	27.62	0.167	0.363	0.151	68.96	0.591	0.192
		Vertical	$R_{M0} = -0.80 - 0.83\cdot \mathbf{ABSQ_{ON}} + 0.08\cdot \mathbf{Polar}$	74.44	21.635	0.194	0.487	0.174	68.02	0.609	0.195
		Normal	$R_{M0} = -1.64 + 0.76\cdot \mathbf{ABSQ_{ON}} + 1.61\cdot \mathbf{Polar}$	72.09	23.289	0.202	0.532	0.182	**72.09**	**0.532**	**0.182**
			DRAGON								
E	5	**Clasic**	$R_{M0} = -1.36 + 0.36\cdot \mathbf{MATS3p} - 0.22\cdot \mathbf{DP11} + 0.09\cdot \mathbf{RDF070e} + 0.37\cdot \mathbf{Mor12u} + 0.20\cdot \mathbf{Ap}$	99.83	1172.9	0.018	0.003	0.014	99.54	0.009	0.023
		Vertical	$R_{M0} = -1.36 + 0.37\cdot \mathbf{MATS3p} - 0.21 * \mathbf{DP11} + 0.08\cdot \mathbf{RDF070e} + 0.38\cdot \mathbf{Mor12u} + 0.20\cdot \mathbf{Ap}$	99.76	794.34	0.021	0.005	0.017	**99.63**	**0.007**	**0.021**
		Normal	$R_{M0} = -1.36 + 0.37 - \mathbf{MATS3p} - 0.21 * \mathbf{DP11} + 0.08\cdot \mathbf{RDF070e} + 0.38\cdot \mathbf{Mor12u} + 0.20\cdot \mathbf{Ap}$	99.76	797.86	0.021	0.005	0.017	**99.76**	**0.005**	**0.017**
F	4	**Clasic**	$R_{M0} = -3.05 - 0.22\cdot \mathbf{DP12} + 0.31\cdot \mathbf{Mor12u} + 0.32\cdot \mathbf{Ap} + 2.07\cdot \mathbf{HATS2u}$	99.06	290.00	0.040	0.018	0.034	98.17	0.035	0.047
		Vertical	$R_{M0} = -3.07 - 0.23\cdot \mathbf{DP12} + 0.32\cdot \mathbf{Mor12u} + 0.33\cdot \mathbf{Ap} + 2.10\cdot \mathbf{HATS2u}$	98.92	248.03	0.043	0.021	0.036	98.03	0.038	0.048
		Normal	$R_{M0} = -3.24 - 0.25\cdot \mathbf{DP12} + 0.30\cdot \mathbf{Mor12u} + 0.34\cdot \mathbf{Ap} + 2.30\cdot \mathbf{HATS2u}$	98.74	225.38	0.047	0.024	0.039	**98.74**	**0.024**	**0.039**
G	3	**Clasic**	$R_{M0} = -3.32 - 0.24\cdot \mathbf{SP13} + 0.32\cdot \mathbf{Ap} + 2.85\cdot \mathbf{HATS2u}$	97.51	156.50	0.063	0.047	0.055	95.56	0.085	0.073
		Vertical	$R_{M0} = -3.40 - 0.19\cdot \mathbf{SP13} + 0.29\cdot \mathbf{Ap} + 2.47\cdot \mathbf{HATS2u}$	96.25	95.009	0.077	0.072	0.067	93.79	0.118	0.086
		Normal	$R_{M0} = -4.37 - 0.26\cdot \mathbf{SP13} + 0.36\cdot \mathbf{Ap} + 3.82\cdot \mathbf{HATS2u}$	94.44	78.309	0.094	0.106	0.081	94.44	0.106	0.081
H	2	**Clasic**	$R_{M0} = -1.64 + 0.61\cdot \mathbf{MATS3v} + 1.61\cdot \mathbf{H1p}$	88.45	49.80	0.130	0.220	0.117	84.97	0.286	0.134
		Vertical	$R_{M0} = -1.63 + 0.76\cdot \mathbf{MATS3v} + 1.59\cdot \mathbf{H1p}$	87.54	53.144	0.135	0.236	0.122	**85.92**	**0.268**	**0.129**
		Normal	$R_{M0} = -1.65 + 0.76\cdot \mathbf{MATS3v} + 1.616\cdot \mathbf{H1p}$	87.41	53.754	0.136	0.240	0.122	**87.41**	**0.240**	**0.122**

6. *Fuzzy* characterization and classification of solvents according to their polarity and selectivity. A comparison with the *Snyder* approach.

6.1. Introduction

The goal of the present study is to describe and apply a new methodology for solvents characterization and classification, improving in this way the approach proposed by *Snyder*. The clustering and classification approaches pro-posed here were iteratively obtained by applying the fuzzy clustering method (partitioning) and robust fuzzy linear discriminant analysis. The diversity of their mathematical backgrounds allows more relevant conclusions to be drawn,finding more specific groups and a better characterization ofsolvents using their degrees of membership to each fuzzy partition and solving in this way some discrepancies.

The efficiency of the robust fuzzy linear discriminantanalysis was measured by the correct classification rate oforiginal data and by the values of quality performance features obtained applying *leave-one-out* (LOO) crossvalidation approach.

6.2. Materials and methods

6.2.1. Dataset

The values of physicochemical characteristics for 72 solvents were taken from the most widely cited classification in chromatography and other separation procedures so called the solvent-selectivity triangle developed by *Snyder* [33]. For this classification, *Snyder* considered different polar interactions as a proton acceptor (x_e), proton donor (x_d), dipole (x_n), chromatographic strength (P') derived from gas-liquid partition coefficient reported by Rohrschneider [34] including also toluene similitudes (x_t) and methylethyl ketone similitudes(x_m). Values of x measure the interaction with test solutes classified as acidic (ethanol), basic (dioxane) and dipolar (nitromethane) and satisfied the following condition: $x_e + x_d + x_n = 1$.

The solvents were assigned to eight groups according to their selectivity. Solvents included in the same group (region) of the triangle have similar selectivity generally whereas from other groups have different selectivity, even if their solventstrength is similar.

6.2.2. Fuzzy clustering methods

To achieve the proposed goal, chemometric methods of multidimensional data analysis were used. The classification methods applied were: the fuzzy c-mean algorithm and fuzzy LDA analysis.

All graphs and chemometric analyzes presented were performed using Statistics 8.1 (StatSoft, Tulsa, USA) and the *Sadic* program.

6.3. Results and discussion

In the following, we tried to resume the *Snyder* procedure, whose classification was based on the physicochemical properties of solvents (selectivity and polarity).

Fuzzy classification algorithms were applied in order to obtain a nuanced, much more natural classification of the 72 solvents used in the study.

6.3.1. Partitional clustering using fuzzy c-means approach

Fuzzy c-means classification (FCM) produced 8 fuzzy partitions (groups), which were all represented by a prototype. To compare the partitions (groups) and the similarity and differences of solvents, we have to analyze both the characteristics of the prototypes corresponding to the eight fuzzy partitions (**A1–A8**) obtained by applying FCM and DOMs of solvents corresponding to all fuzzy partitions.

The fuzzy partition **A1**, for example, has a moderatechromatographic strength (P' = 4.71) and proton acceptor ($\tilde{x}_e$ = 0.23), but a high dipolarity ($\tilde{x}_d$ = 0.23) and very small proton donor ($\tilde{x}_d$ = 0.23). This partition includes all the solvents ofthe Snyder's group VIa excepting formyl morpholine and butyrolactone which are assigned to the partition **A4** with avery high DOM (0.9841) and moderate one (0.7732), respectively. To this partition (**A1**) were assigned in addition3 solvents from Snyder group III (2-picoline, 2,6-lutidine, quinoline) with moderate DOMs, and nitrobenzene from Snyder group VII with a relatively small DOM (0.5656).

The fuzzy partition **A2**, with relatively high value for P' (7.09), $\tilde{x}_e$ (0.39) and $\tilde{x}_n$ (0.35), contain different solvents including dimethyl sulfoxide (III), ethyleneglycol (IV) and m-crezol (VIII) with the following very high DOMs: 0.9679 > 0.8757 > 0.8324.

By the contrary, the fuzzy partition **A3** including the majority of solvents assigned by Snyder to the group I (aliphatic ethers and triethylamine) and VII (benzene and its derivatives) has the smallest chromatographic strength (P' = 2.48) and a high dipolarity ($\tilde{x}_n$ = 0.42).

The fuzzy partition **A4** has a high chromatographic strength (P' = 6.25), proton acceptor ($\tilde{x}_e$ = 0.37) and dipolarity ($\tilde{x}_n$ = 0.38), but a moderate proton donor ($\tilde{x}_d$ = 0.25). This partition contains solvents of the Snyder's group III (dimethylformamide, *N,N*-dimethyl acetamide, methylformamide, tetramethylurea), and also aniline (VIb), nitromethane (VII), acetic acid (IV), but with quite different DOM: 0.9530 > 0.7272 > 0.7259.

The fuzzy partition **A5** has a medium chromatographic strength (P' = 3.99), but the highest proton acceptor value ($\tilde{x}_e$ = 0.47). This partition includes the majority of aliphatic alcohols (Snyder' group II) with high DOMs and also tetrahydrofuran (Snyder' group III) with a high DOM = 0.8888 and chloroform (Snyder' group VIII) with a moderate DOM = 0.6346.

The fuzzy partition **A6** has a small chromatographic strength (P' = 3.28), but the highest dipolarity ($\tilde{x}_n$ = 0.44). This partition includes solvents from Snyder's group VII with very high DOMs (ethoxybenzene, fluorobenzene, diphenylether) and also methylene chloride, ethylene chloride (Snyder' group V) with high DOMs and 1-Octanol (Snyder' group II) with a small DOM = 0.5877.

The fuzzy partition **A7** has a moderate chromatographic strength (P' = 5.43), and high proton acceptor value ($\tilde{x}_e$ = 0.38) and dipolarity ($\tilde{x}_n$ = 0.37). This partition includes solvents from Snyder' group III with very high DOMs (nonylphenol oxyethylate, 2-methoxyethanol, pyridine, triethyleneglycol, diethyleneglycol), and also benzyl acohol (IV), nitroethane (VII), acetonitrile (VIb), methanol (II), acetone (VIa) with DOMs in the following order: 0.6972 > 0.6814 > 0.5008 > 0.4954 > 0.4942.

The fuzzy partition **A8** has the highest chromatographic strength (P' = 9.44) and methylethyl ketone similitudes ($\tilde{x}_m$ = 0.35), but the smallest dipole ($\tilde{x}_n$ = 0.28). Also, quite interesting, the proton acceptor value ($\tilde{x}_e$ = 0.35) and donor acceptor value ($\tilde{x}_d$ = 0.36) are near equal. This group is identical to the Snyder's group VIII, excepting formamide which is assigned also to this group of water with a very high DOM (0.9866). Chloroform assigned to this group in Snyder classification was moved in fuzzy partition **A5** (a group of alcohols).

6.3.2. Fuzzy discriminant analysis (FLDA)

The application of FLDA led to the discrimination of the investigated solvents in eight classes in good agreement to classification of *Snyder* but with very various and relatively small DOMs, from 0.2251 for formamide and 0.9110 for tert-pentanol (**Table 6.2**). The results indicate a very good separation of solvents in about all cases according to the correct classification rate (**Table 6.3**): the highest value (100%) was obtained in the majority of cases (II, IV, V and VI) and the lowest value (66.67%) for solvents corresponding to the first group.

Tabel 6.3. Classification matrix of solvents obtained by applying fuzzy discriminant analysis for the eight groups of Snyde.

		Classification matrix								Classification matrix (%)							
Class	**Total**	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**
I	**6**	**4**	0	**2**	0	0	0	0	0	**66.67**	0	**33.33**	0	0	0	0	0
II	**8**	0	**8**	0	0	0	0	0	0	0	**100**	0	0	0	0	0	0
III	**15**	0	0	**13**	0	0	0	0	0	0	0	**86.67**	0	0	**6.67**	**6.67**	0
IV	**4**	0	0	0	**4**	0	0	0	0	0	0	0	**100**	0	0	0	0
V	**2**	0	0	0	0	**2**	0	0	0	0	0	0	0	**100**	0	0	0
VI	**18**	0	0	0	0	0	**18**	0	0	0	0	0	0	0	**100**	0	0
VII	**14**	0	0	0	0	0	**2**	**12**	0	0	0	0	0	0	**14.29**	**85.71**	0
VIII	**5**	0	0	0	0	0	0	**1**	**4**	0	0	0	0	0	0	**20.00**	**80.00**

In the case of first group, two solvents hexamethyl phosphoramide and tetramethyl guanidine, were moved to group III according to the following small DOMs: 0.3563 and 0.4524. By the other side, two solvents from group III, namely tetrahydrofuran and nonylphenol oxyethylate were moved to other groups: tetrahydrofuran to VI (DOM = 0.3933) and nonylphenol oxyethylate to VII (DOM = 0.2265). Two solvents included by *Snyder* to the group VII were assigned to group VI: nitromethane (0.4171) and nitroethane (0.4223). The case of group VIII is also quite interesting because chloroform included in this group by *Snyder* is move in the group VII (DOM = 0.2693).

The efficiency of FLDA are well supported by scatterplot of canonical scores on the plane defined by ROOT1-ROOT2(**Figure 6.2**).

Table 6.2. Results obtained for eight predefined groups applying fuzzy linear discriminant analysis.

Class	Parameters of class centers						Solvents (ranked in decreasing order)	Name of solvents	DOMs
	P'	$\tilde{x}_e$	$\tilde{x}_d$	$\tilde{x}_n$	$\tilde{x}_t$	$\tilde{x}_m$			
I	2.4740	0.5003	0.1438	0.3557	0.0816	0.1785	23 22 41 24	Di-*iso*-propiletilenă (I) Dietileter (I) Trietilamină (I) Di-*n*-butileter (I)	0.7878 0.7764 0.5212 0.4742
II	4.0083	0.5481	0.1927	0.2656	0.0531	0.2118	47 3 7 4 5 2	*terţ*-Pentanol (II) 1-Propanol (II) 1-Octanol (II) 2-Propanol (II) 1-Butanol (II) Etanol (II)	0.9110 0.9064 0.8361 0.7903 0.7852 0.5646
III	5.6099	0.4124	0.2191	0.3692	0.1234	0.2353	43 61 44 52 71 39 48 57 38 50 37 55 65	Piridină (III) Metilformamidă (III) Quinolină (III) 2-Picolină (III) Trietilenglicol Tetrametiluree (III) 2-Metoxietanol (III) *N,N*-dimetil acetamidă (III) *N*-Metilpirolidină (III) 2,6-lutidină (III) Dimetilformamidă (III) Tetrametil guanidină (I) Dimetil sulfoxid (III)	0.9063 0.8295 0.6801 0.6651 0.5697 0.5676 0.5652 0.5462 0.4867 0.4851 0.4754 0.4524 0.4459
IV	6.2920	0.4046	0.3013	0.2938	0.1084	0.2752	9 30 8	Benzil alcool (IV) Acid acetic (IV) Etilenеglicol (IV)	0.8592 0.7619 0.6645
V	3.3000	0.2950	0.1950	0.5100	0.2000	0.2100	49	Clorură de metilen (V)	0.4132

VI	5.3804	0.3389	0.2507	0.4112	0.1432	0.2413	62 33 35 27 21 68 32 36 20 31 58 69 54 67 64 19 18	Ciano morfolină (VI)	0.7833
								Acetofenonă (VI)	0.7346
								Carbonat de propilen (VI)	0.7136
								Dioxan (VI)	0.7112
								Acetonitril (VI)	0.6543
								tris-cianoetoxipropan (VI)	0.6260
								Ciclohexanonă (VI)	0.6201
								Butirolactonă (VI)	0.6035
								Benzonitril (VI)	0.5813
								Acetonă (VI)	0.5531
								Tricrezilfosfat (VI)	0.5489
								Oxidipropionitril (VI)	0.5456
								bis-(2-etoxi etil) eter (VI)	0.5267
								Tetrahidrotiofenă (VI)	0.5232
								Formil morfolină (VI)	0.4554
								Nitroetan (VII)	0.4223
								Nitrometan (VII)	0.4171
VII	3.2526	0.2573	0.3072	0.4354	0.1554	0.3142	45 13 16 12 14 51 28 26 15 17 25	Fluorobenzen (VII)	0.8783
								Clorobenzen (VII)	0.8110
								p-Xilen (VII)	0.8108
								Benzen (VII)	0.8028
								Bromobenzen (VII)	0.7974
								Etoxibenzen (VII)	0.7696
								Anisol (VII)	0.7525
								Difenileter (VII)	0.7058
								Iodobenzen (VII)	0.6423
								Nitrobenzen (VII)	0.5931
								Dibenzileter (VII)	0.4128
VIII	8.5477	0.3486	0.3768	0.2727	0.0772	0.3617	63 66 72 56	Dodecafluoroheptanol (VIII)	0.8983
								Tetrafluoropropanol (VIII)	0.7597
								Apa (VIII)	0.6239
								m-Crezol (VIII)	0.5920

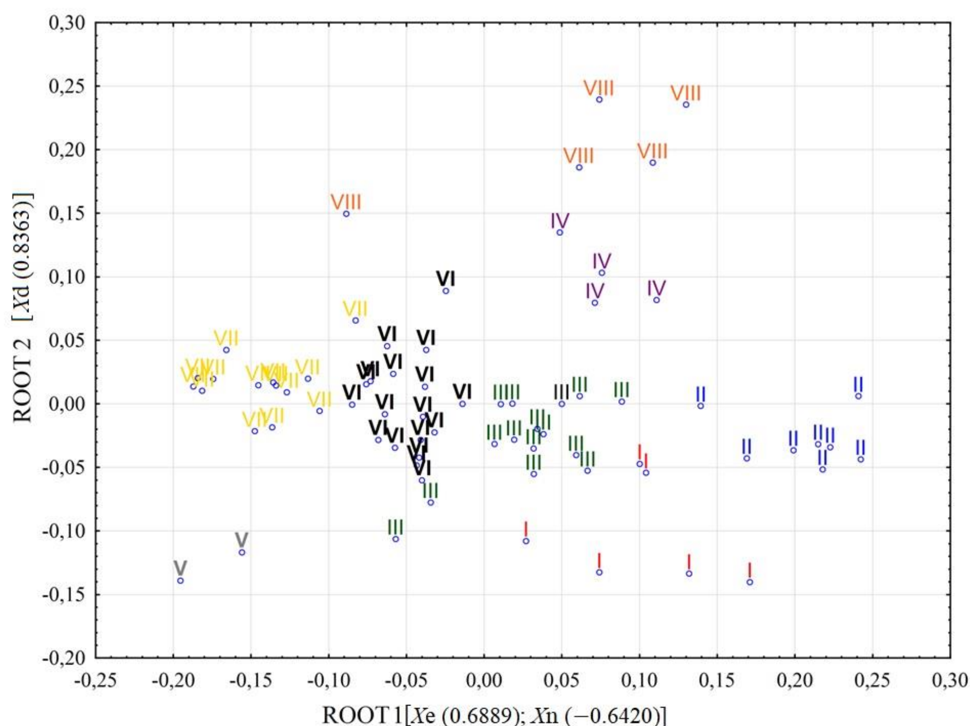


Figure 6.2. Scatterplot of canonical scores corresponding to ROOT1 and ROOT2.

At the same time, the results of the classification matrix using the *leave-one-out* cross-validation procedure (**Table 6.4**), showed a high correctness of the classification rate in accordance with the classification of the initial data.

Table 6.4. Classification matrix of solvents obtained by applying *leave-one-out* cross-validation approach.

		Classification matrix								Classification matrix (%)							
Class	**Total**	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**
I	**6**	**4**	0	**2**	0	0	0	0	0	**66.67**	0	**33.33**	0	0	0	0	0
II	**8**	0	**8**	0	0	0	0	0	0	0	**100**	0	0	0	0	0	0
III	**15**	0	0	**13**	0	0	0	0	0	0	0	**86.67**	0	0	**6.67**	**6.67**	0
IV	**4**	0	0	0	**3**	0	0	0	0	0	0	0	**75.00**	0	0	0	**25.00**
V	**2**	0	0	0	0	**2**	0	0	0	0	0	0	0	**100**	0	0	0
VI	**18**	0	0	0	0	0	**17**	0	0	0	0	0	**5.56**	0	**94.44**	0	0
VII	**14**	0	0	0	0	0	**2**	**12**	0	0	0	0	0	0	**14.29**	**85.71**	0
VIII	**5**	0	0	0	0	0	0	**1**	**4**	0	0	0	0	0	0	**20.00**	**80.00**

7. Characterization and classification of foods based on amino acids content, using fuzzy c-means algorithms

7.1. Introduction

The aim of this study was to characterize and classify the most important and consumed foods according to their AA profile, especially according to their EAA content, using advanced chemometric techniques.

7.2. Materials and methods

7.2.1. Samples

The data analyzed in this study include 100 samples (food) characterized by the content of the most important AA, collected from various sources in the literature [35−43].

The investigated foods are classified in: milk, eggs, meat, fruits, vegetables, nuts, seeds, honey and wine and include information on the content in AA.

AA analyzed by different analytical methods (RP-HPLC, IEC, GC) for all considered foods were the following : alanine (**Ala**), arginine (**Arg**), aspartic acid (**Asp**), cysteine (**Cys**), glutamic acid (**Glu**), glycine (**Gly**), histidine (**His**), isoleucine (**Ile**), leucine (**Leu**), lysine (**Lys**), methionine (**Met**), phenilalanine (**Phe**), proline (**Pro**), serine (**Ser**), threonine (**Thr**), triptophan (**Trp**), tyrozine (**Tyr**), valine (**Val**). The content of amino acid for each sample (food) ware expressed in percentage (%) of the mass, resulting a matrix with the size of 100×18.

7.2.2. Methods

To achieve the proposed goal, were used the classical HCA analysis and metoda fuzzy divisive c-means algorithm.

7.3. Results and disscusion

The data collected on the AA content of the most consumed foods is presented in **Figure 7.1**. As can be seen, the most abundant AA are Pro, Glu and Asp. With the exception of three AAs, most foods have an AA content ranging from 0−10%

Considering the obtained dendrogram (**Figure 7.2**), applying HCA analysis (complete linkage and Euclidian distance as a measure of similarity) a very important aspect can be observed, especially that the difference between foods increases exponentialy from right to left, except for the group of fish that are very similar to each other, but also clearly different from the other foods.

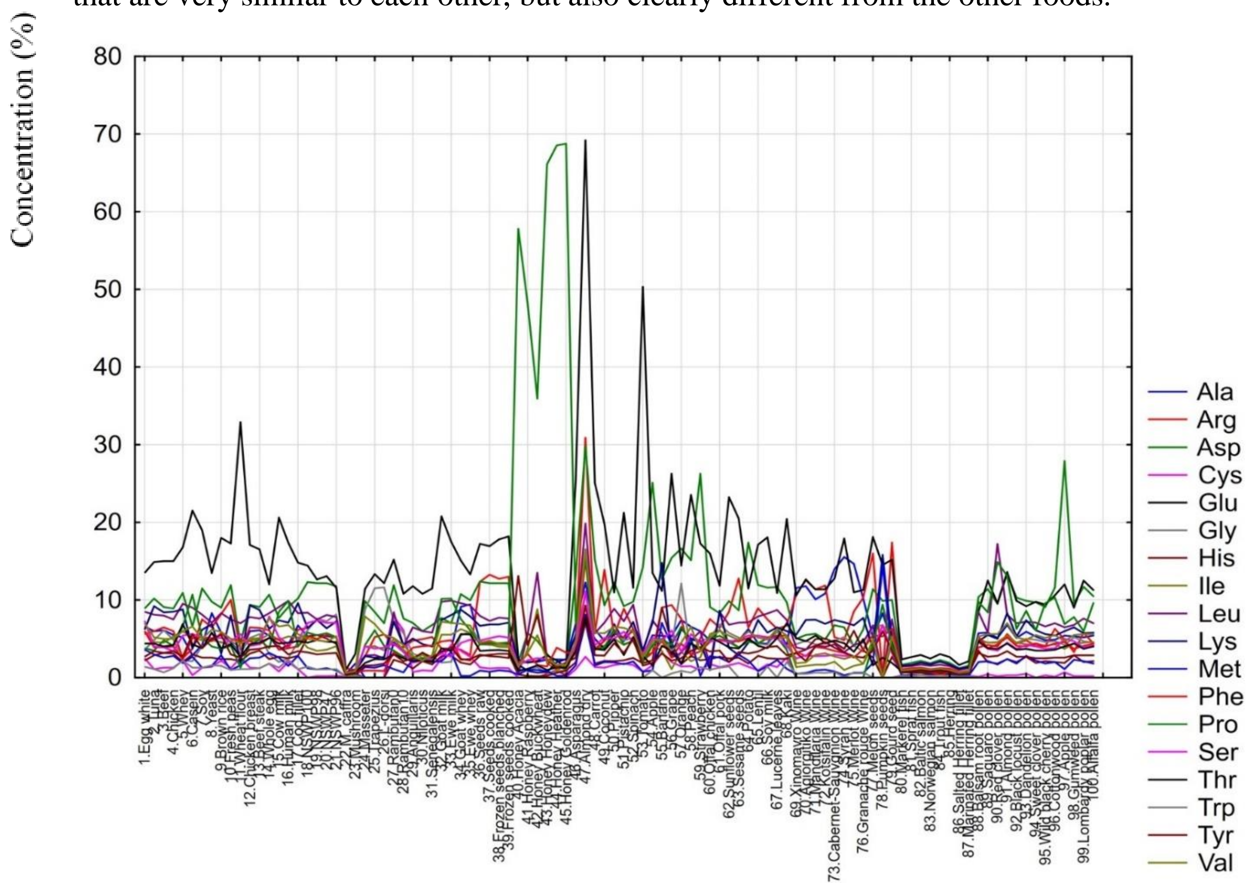


Figure 7.1. Profile of food samples in terms of amino acid content .

To compare the partitions obtained by applying the fuzzy c-means approch, we analyzed the similarities and differences between foods, the degrees of membership (DOMs) corresponding to all

fuzzy partitions for both samples and characteristics (AA concentration). The results obtained applying the method of hierarchical-divisive associative clusters are presented in **Table 7.1**. Through a rigorous analysis of the fuzzy partitions at each level (partition history) in parallel with the values of AA concentrations in food, the following remarks can be made. At the first partition level (100 samples) they were separated into 2 fuzzy partitions **A1** and **A2**. The DOMs values of the foods in partition **A1** are in the range of 0.029−0.945 or 2.90−94.50% and between 0.962−0.994 (99.62−99.40%) in the case of fuzzy partition **A2**. Most of the analyzed foods have been assigned to the fuzzy partition **A1**.

Table 7.1. Partițiile fuzzy asociate finale.

Fuzzy partition level	**Fuzzy divisive partition**	**Samples**	**DOMs**	**Associate variables** (concentration of AA, %)	**DOMs**
0	A	1,…, 100 Foods		1,..., 18	
Final partition					
1	**A111111**	57, 47	**0.501; 0.029**	**Gly**	**0.821**
2	**A111112**	20, 19, 21, 18 (**7.05, 7.36, 7.07, 7.09**)	**0.918−0.590**	**Ser** (highest)	**0.669**
3	**A11112**	35, 34 (**Val: 6.9, 6.7**; Ala: 9.1, 9.4)	**0.881;0.704**	**Val**, **Ala**	**0.722; 0.446**
4	**A11121**	14, 52, 67, 50, 61, 1, 68	**0.764−0.408**	**Thr**	**0.709**
5	**A11122**	64, 9, 91, 8, 88 (***Phe***: ***5.1, 5.7, 5.2, 4.7, 5.3***)	**0.585−0.260**	**Phe**; **Ile** (**mare**)	**0.854; 0.792**
6	**A1121**	49, 37, 77, 36, 38, 63, 39, 79, 72, 71, 10, 51, 62	**0.901−0.389**	**Arg** (highest) ~ **12**	**0.624**
7	**A11221**	2, 3, 4, 12, 13, 17, 5, 93, 65, 73, 78, 55	**0.945−0.269**	**Lys**	**0.753**
8	**A11222**	16, 15, 32, 33, 90, 7, 66, 27, 6	**0.739−0.262**	**Leu**	**0.510**
9	**A121**	46, 95, 92, 89, 59, 54, 48, 42, 97, 94, 76, 69, 96, 74, 41, 99, 100, 45, 25, 11, 70, 43, 44, 53, 30, 26, 29, 98, 31, 28, 24	**0.892−0.363**	**His**; **Met** **Tyr**	**0.867; 0.792 0.486**

10	**A1221**	40, 75	**0.501; 0.363**	**Trp**	**0.803**
11	**A1222**	56, 58, 60	**0.738−0.628**	**Cys**	**0.877**
12	**A2**	82, 84, 81, 85, 83, 80, 23, 87, 86, 22	**0.994−0.962**	**Asp, Glu** (lower), **Pro** (lower)	**0.990; 0.879 0.754**

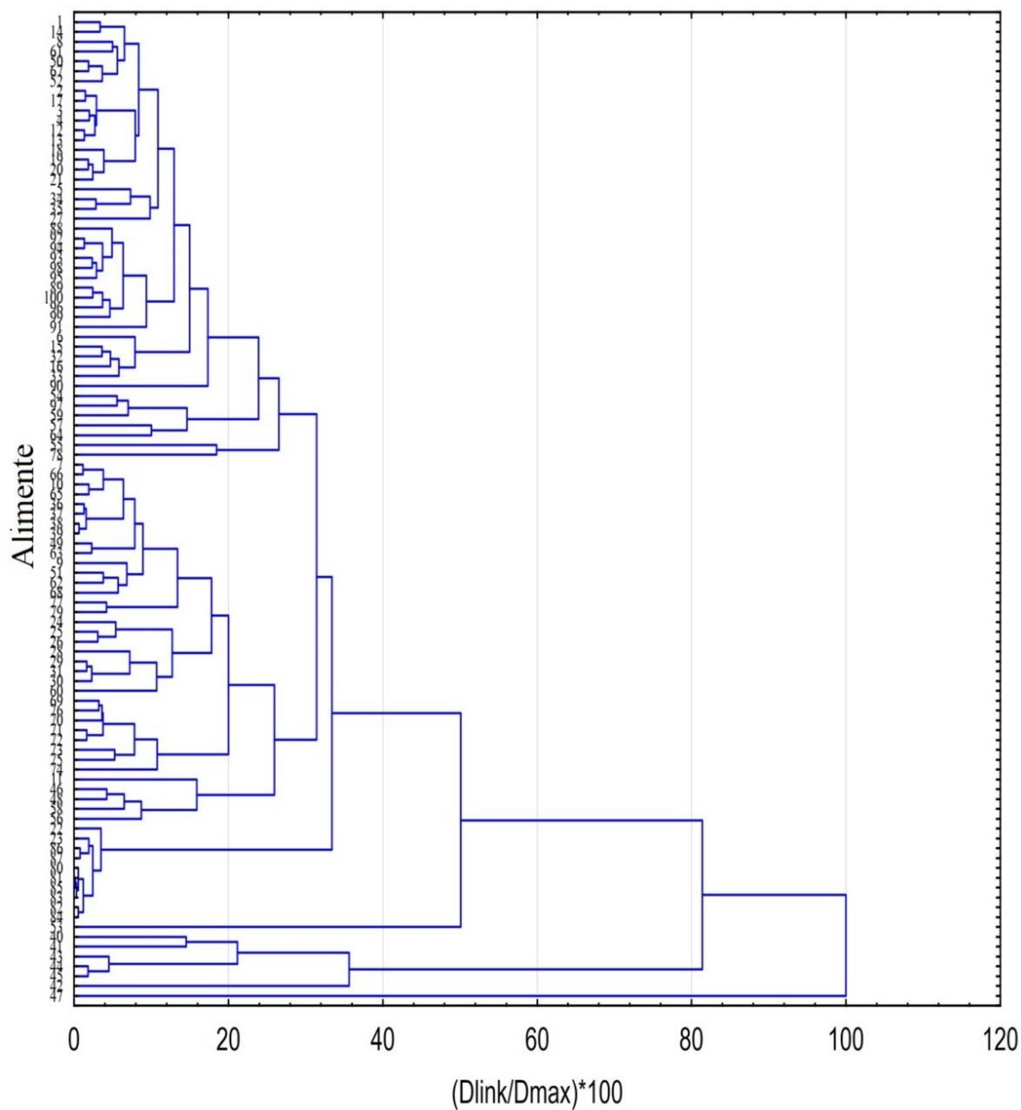


Figure 7.2. Dendogram corresponding to all investigated food samples(18 amino acids).

CHAPTER IV—GENERAL CONCLUSIONS

- Applying various experimental methods and advanced chemometric analysis, we could characterize and classify amino acids and related compounds according to their specific properties (lipophilicity and antioxidant activity).
- Proteinogenic AAs were tested using different methods for antioxidant activity determination based on several types of free radicals (present both *in vivo* and *in vitro*). The results of the study confirmed that some proteinogenic AAs have a significant free radical scavenging activity through most applied methods and support the development of a new class of multifunctional and natural antioxidants. The Amino acids with high activity are: **Cys**, **Trp**, **Tyr**, **Arg** and **Asn**.
- PCA and CA analysis showed that antioxidant activity methods based on similar mechanisms are closely related. Based on SRD results, the methods that best discriminate AA are those based on the scavenging antioxidant power, such as DPPH and NO.
- Regarding catecholamines, based on the obtained results, they shown a significant activity of scavenging and reducing the power of free radicals and support the development of a new class of multifunctional antioxidants.
- According to the experimental and chemometric results, most drugs have high antioxidant activity in most assays, except for methoxamine, L-dopa and D-dopa (which do not have pharmacological properties). At the same time, the investigated compounds showed a low chelating capacity of ferrous ions.
- Regarding the prediction and modeling of lipophilicity of AAs, the best models, validated internally by cross-validation procedure, showed that a small number of molecular descriptors are needed to obtain statistically significant prediction models.
- Models derived from Dragon descriptors showed greater predictive power compared to those derived from Alchemy, because they describe much better the 3D aspects of AAs.
- The selected descriptors in the equations of the most powerfull models for prediction of the lipophilicity of the investigated compounds contain both 2D and 3D aspects of the molecular structure, as well as aspects regarding the topology, conformation, connectivity indices and some molecular properties of the compounds.

- Fuzzy models compared to the classic ones, demonstrated a much higher predictive power for both models containing Dragon and Alchemy descriptors, and also, presenting a much better and efficient applicability for prediction of the molecular property with great interest in drug discovery, toxicology and ecology.
- By applying robust fuzzy classification algorithms, it was possible to find more specific groups and better characterization of solvents using the degree of membership for each fuzzy partition and thus solving the problem of solvents that were not well classified according to *Snyder*'s classification.
- At the same time, we managed to highlight more efficiently the specific characteristics of each class of solvents, in terms of prototype parameters and class centers. Thus, by applying fuzzy algorithms, the solvent or mixture of solvents suitable for the desired chromatographic analysis could be selected much more easily and correctly.
- By applying fuzzy c-means algorithm, it was possible to classify foods in terms of AA content, especially in EAA content. With the help of prototype parameters and class centers, the specific characteristics of each food class were highlighted much more efficiently and well, and the degrees of membership allowed a much more rational comparison of (di)similarities between foods according to their AA content.
- The applied advance chemometric analysis proved, once again, that they are extremely useful for studying the similarities and differences between foods, on the one hand, and AA on the other hand, as well as establishing the AA characteristics for each food class.

Selective bibliography

(1) Davis, T. A.; Fiorotto, M. L. Regulation of Muscle Growth in Neonates. *Curr. Opin. Clin. Nutr. Metab. Care,* **2009**, *12 (1)*, 78–85. https://doi.org/10.1097/MCO.0b013e32831cef9f.

(2) Hardy, P. M. The Protein Amino Acids. In *Chemistry and Biochemistry of the Amino Acids*; Springer: Netherlands, 1985; pp 6–24. https://doi.org/10.1007/978-94-009-4832-7_2.

(3) Wu, G. *Amino Acids: Biochemistry and Nutrition*. CRC Press: New York, 2010.

(4) Li, P.; Yin, Y. L.; Li, D.; Kim, W. S.; Wu, G. Amino Acids and Immune Function. *Br. J. Nutr*. **2007**, *98 (2)*, 237–252. https://doi.org/10.1017/S000711450769936X.

(5) Amit, Kessel; Nir, B.-T. Introduction to Proteins: Structure, Function, and Motion. In *CRC Press*; CRC Press: Boca Raton, FL, USA, 2018; p 932.

(6) Harold E. Himwich; Williamina A. Himwich. *Biogenic Amines*; Elsevier: Amsterdam, 1964; Vol. 8.

(7) Nagatsu, T.; Nabeshima, T.; McCarty, R.; Goldstein, D.S. *Catecholamine Research: From Molecular Insights to Clinical Medicine*; Kluwer Academic/Plenum Publisher: New York, 2002. https://doi.org/10.1007/978-1-4757-3538-3.

(8) Seiji, S.; Ikuo, H.; Nobumasa, K. *Transmitters and Modulators in Health and Disease: New Frontiers in Neuroscience*; Springer: Tokyo, Japan, 2009.

(9) Medina, M. Á.; Urdiales, J. L.; Rodríguez-Caso, C.; Ramírez, F. J.; Sánchez-Jiménez, F. Biogenic Amines and Polyamines: Similar Biochemistry for Different Physiological Missions and Biomedical Applications. *Crit. Rev. Biochem. Mol. Biol.* **2003**, *38* (1), 23–59. https://doi.org/10.1080/713609209.

(10) Mans, B. J.; Ribeiro, J. M. C.; Andersen, J. F. Structure, Function, and Evolution of Biogenic Amine-Binding Proteins in Soft Ticks. *J. Biol. Chem.* **2008**, *283* (*27*), 18721–18733. https://doi.org/10.1074/jbc.M800188200.

(11) Miura, T.; Muraoka, S.; Ogiso, T. Antioxidant Activity of Adrenergic Agents Derived from Catechol. *Biochem. Pharmacol.* **1998**, *55* (*12*), 2001–2006. https://doi.org/10.1016/S0006-2952(98)00075-6.

(12) Kawashima, T.; Ohkubo, K.; Fukuzumi, S. Radical Scavenging Reactivity of Catecholamine Neurotransmitters and the Inhibition Effect for DNA Cleavage. *J. Phys. Chem. B* **2010**, *114* (*1*), 675–680. https://doi.org/10.1021/jp909314t.

(13) Sitole, L. J.; Tugizimana, F.; Meyer, D. Multi-Platform Metabonomics Unravel Amino Acids as Markers of HIV/Combination Antiretroviral Therapy-Induced Oxidative Stress. *J. Pharm. Biomed. Anal.* **2019**, *176*, 112796. https://doi.org/10.1016/j.jpba.2019.112796.

(14) Bourdon, E.; Blache, D. The Importance of Proteins in Defense against Oxidation. *Antioxidants Redox Signal.* **2001**, *3* (*2*), 293–311. https://doi.org/10.1089/152308601300185241.

(15) Kim, J. H.; Jang, H. J.; Cho, W. Y.; Yeon, S. J.; Lee, C. H. In Vitro Antioxidant Actions of Sulfur-Containing Amino Acids. *Arab. J. Chem.* **2018**, *13* (*1*), 1678–1684. https://doi.org/10.1016/j.arabjc.2017.12.036.

(16) Atmaca, G. Antioxidant Effects of Sulfur-Containing Amino Acids. *Yonsei Med. J.* **2004**, *45* (*5*), 776. https://doi.org/10.3349/ymj.2004.45.5.776.

(17) Meucci, E.; Mele, M. C. Amino Acids and Plasma Antioxidant Capacity. *Amino Acids* **1997**, *12* (*3–4*), 373–377. https://doi.org/10.1007/BF01373017.

(18) Sila, A.; Bougatef, A. Antioxidant Peptides from Marine By-Products: Isolation, Identification and Application in Food Systems. A Review. *J. Funct. Foods* **2016**, *21*, 10–26. https://doi.org/10.1016/j.jff.2015.11.007.

(19) Yang, X.-R.; Qiu, Y.-T.; Zhao, Y.-Q.; Chi, C.-F.; Wang, B. Purification and Characterization of Antioxidant Peptides Derived from Protein Hydrolysate of the Marine Bivalve Mollusk Tergillarca Granosa. *Mar. Drugs* **2019**, *17* (*5*), 251. https://doi.org/10.3390/md17050251.

(20) Guo, H.; Ekusa, A.; Iwai, K.; Yonekura, M.; Takahata, Y.; Morimatsu, F. Royal Jelly Peptides Inhibit Lipid Peroxidation In Vitro and In Vivo. *J. Nutr. Sci. Vitaminol. (Tokyo).* **2008**, *54* (*3*), 191–195. https://doi.org/10.3177/jnsv.54.191.

(21) Dash, P.; Ghosh, G. Amino Acid Composition, Antioxidant and Functional Properties of Protein Hydrolysates from Cucurbitaceae Seeds. *J. Food Sci. Technol.* **2017**, *54* (*13*), 4162–4172. https://doi.org/10.1007/s13197-017-2855-6.

(22) Chanin Nantasenamat; Chartchalerm Isarankura-Na-Ayudhya; Thanakorn Naenna; Virapong Prachayasittikul; Virapong Prachayasittikul. A Practical Overview of Quantitative Structure-Activity Relationship. *EXCLI J.* **2009**, *8*, 74–88.

(23) Nantasenamat, C.; Isarankura-Na-Ayudhya, C.; Prachayasittikul, V. Advances in Computational Methods to Predict the Biological Activity of Compounds. *Expert Opin. Drug Discov.* **2010**, *5* (*7*), 633–654. https://doi.org/10.1517/17460441.2010.492827.

(24) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; *et. al.* A. QSAR Modeling: Where Have You Been? Where Are You Going To? *J. Med. Chem.* **2014**, *57* (*12*), 4977–5010. https://doi.org/10.1021/jm4004285.

(25) Prachayasittikul, V.; Pingaew, R.; Anuwongcharoen, N.; Worachartcheewan, A.; Nantasenamat, C.; Prachayasittikul, S.; Ruchirawat, S.; Prachayasittikul, V. Discovery of Novel 1,2,3-Triazole Derivatives as Anticancer Agents Using QSAR and in Silico Structural Modification. *Springerplus* **2015**, *4* (*1*), 571. https://doi.org/10.1186/s40064-015-1352-5.

(26) Katritzky, A.; Ramsden, C.; Joule, J.; Zhdankin, V. *Handbook of Heterocyclic Chemistry* , 3rd edition.; Elsevier: Amsterdam, 2010.

(27) Leo, A.; Hansch, C.; Elkins, D. Partition Coefficients and Their Uses. *Chem. Rev.* **1971**, *71* (*6*), 525–616. https://doi.org/10.1021/cr60274a001.

(28) James M. Sangster. *Octanol-Water Partition Coefficients: Fundamentals and Physical Chemistry*; Wiley, 1997.

(29) Van de Waterbeemd, H.; Kansy, M.; Wagner, B.; Fischer, H. Lipophilicity Measurement by High-Performance Liquid Chromatography (RP-HPLC). In *Lipophilicity in drug action and toxicology*; Pilska, V.; Testa, B.; Van de Waterbeemd, H. Eds.; VCH: Weinheim, 1996. https://doi.org/10.1002/9783527614998.

(30) Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M. MobyDigs: Software for Regression and Classification Models by Genetic Algorithms. In *Data Handling in Science and Technology*; Elsevier, **2003**; Vol. 23, pp 141–167.

https://doi.org/10.1016/S0922-3487(03)23005-7.

(31) Mercader, A. G.; Duchowicz, P. R.; Fernández, F. M.; Castro, E. A. Modified and Enhanced Replacement Method for the Selection of Molecular Descriptors in QSAR and QSPR Theories. *Chemom. Intell. Lab. Syst.* **2008**, *92* (*2*), 138–144. https://doi.org/10.1016/j.chemolab.2008.02.005.

(32) Casoni, D.; Sârbu, C. Comparative Study of Different Oils and Fats Impregnated Thin-Layer Chromatographic Layers for Amino Acids Lipophilicity Estimation. *Stud. UBB Chem.* **2011**, *LVI* (*1*), 45–61.

(33) Snyder, L. R. Classification off the Solvent Properties of Common Liquids. *J. Chromatogr. Sci.* **1978**, *16* (*6*), 223–234. https://doi.org/10.1093/chromsci/16.6.223.

(34) Rohrschneider, L. Solvent Characterization by Gas-Liquid Partition Coefficients of Selected Solutes. *Anal. Chem.* **1973**, *45* (*7*), 1241–1247. https://doi.org/10.1021/ac60329a023.

(35) Amino Acid Content of Food and Biological Data on Proteins http://www.fao.org/3/AC854T/AC854T03.htm#chI.I (acesat în 5 Februarie 2021)

(36) Mehdizadeh, S.; Lasekan, O.; Muhammad, K.; Baharin, B. Variability in The Fermentation Index, Polyphenols and Amino Acids of Seeds of Rambutan (Nephelium Lappaceum L.) during Fermentation. *J. Food Compos. Anal.* **2015**, *37*, 128−135

(37) Adeyeye, E.I. Amino acid composition of three species of Nigerian fish: Clarias anguillaris, Oreochromis niloticus and Cynoglossus senegalensis. *Food Chem.* **2009**, *113*, 43−46

(38) Hejtmankova, A.; Pivec, V.; Trnkova, E.; Dragounova, H. Differences in The Composition of Total and Whey Proteins in Goat and Ewe Milk and Their Changes Throughout The Lactation Period. *Czech. J. Anim*. Sci. **2012**, *57*, 323−331

(39) Lisiewska, Z.; Kmiecik, W.; Slupski, J. Content Of Amino Acids in Raw and Frozen Broad Beans (Vicia Faba Var. Major) Seeds At Milk Maturity Stage, Depending On The Processing Method. *Food Chem.* **2007**, *105*, 1468−1473.

(40) Janiszewska, K.; Aniolowska, M.; Nowakowski, P. Free Amino Acids Content of Honeys from Poland. *Pol. J. Food Nutr. Sci.* **2012**, *62*, 85−89.

(41) Bouloumpasi, E.; Soufleros, E.H.; Tsarchopoulos, C.; Biliaderis, C.G. Primary amino acid composition and its use in discrimination of Greek red wines with regard to variety and cultivation region. *Vitis*, **2002**, *41* (4), 195−200

(42) Adeyemi, F.O.; Adediran, G.O. Amino Acid and Mineral Compositions and Functional Properties of Some Oilseeds. J. Agric. *Food Chem.* **1994**, *42*, 878−881.

(43) Mccaughey, WF; Gilliam, M; Standifer. Amino acids and protein adequacy for honey bees of pollens from desert plants and other floral sources. *Apidologie*, **1980**, *11*(1), 75−86