Quantum-Chemical Study of the Biogenic Amino Acids

Roman Boča (roman.boca@ucm.sk)
Univerzita sv Cyrila a Metoda v Trnave
Beata Vranovičová
Univerzita sv Cyrila a Metoda v Trnave

Research Article

Keywords: amino acids, ab initio calculations, PM3 calculations, molecular properties, multivariate methods

Posted Date: April 12th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-393097/v1

Abstract

Ten amino acids have been subjected to the quantum chemical calculations using the ab initio MO-LCAO-SCF calculations and semiempirical PM3 method. When the geometry optimization started form the X-ray structure confirming the zwitterionic form, the ab initio calculations in vacuo result in the amino acid (canonical) form with the hydrogen atom attached not to the amine but to the carboxylate group. At the optimum geometry a number of properties were evaluated: dipole moment, dipole polarizability, molecular surface, molecular volume, HOMO, LUMO, ionization energy and electron affinity using the ΔSCF approach and their values corrected for electron correlation by the 2 nd –order perturbation theory (MP2). In addition, the molecular electrostatic potential and the charge density have been drawn. These properties have been mutually correlated by employing the statistical multivariate methods: the cluster analysis, the probabilistic neural network classifier, the principal component analysis and the Pearson pair correlation.

1 Introduction

Amino acids occur as building blocks of proteins and other polypeptides. Of amino acids that make up proteins (biogenic amino acids), 10 were selected for the present theoretical study that follows up previous voltametric experiments. According to the ability be synthesized in organisms, they can be classified as non-essential (alanine Ala, asparagine Asn, glutamic acid Glu), conditionally essential (arginine Arg, cysteine Cys, glycine Gly, tyrosine Tyr) and essential (histidine His, phenylalanine Phe, tryptophan Trp) [1]. They can be classified also by their electric charge at side chain (Arg, His, Glu), polar uncharged side chain (Asn), hydrophobic side chain (Ala, Phe, Tyr, Trp), or special cases (Gly, Cys) [2,3].

Theoretical investigations of aminoacids by contemporary quantum chemical methods confirmed that involvement of several water molecules into the model results the stabilization of the canonical (aminoacid) form preferred in vacuo to the zwitterionic form as observed in solutions [4,5].

Amino acids can either be oxidized or reduced as confirmed by the cyclic voltammetry. The oxidation and/or reduction potentials depend upon pH of the sampled solution [6,7]. These species manifest their redox properties during their spontaneous reactions with iron(II) salts [8,9]: under anaerobic conditions they are capable of liberating colloidal Fe(0) from Fe(II) salts. Thus their reducing ability is associated with their oxidation potential.

2 Methods

The quantum-chemical calculations have been done using the MO-LCAO-SCF approach in two versions [10]: (i) a semiempirical PM3 method is based upon the parametrization and a zero-differential overlap approximation; this is fast but less reliable; (ii) *ab initio* calculations were conducted with the STO 6-31** basis set functions. In both cases the full geometry optimization has been performed starting from the geometry configuration as retrieved from the Cambridge Crystallographic Data Centre [11]. In the

optimum geometry the set of scalar molecular properties has been evaluated: the dipole moment, dipole polarizability, molecular surface, molecular volume, energies of the HOMO and LUMO pair. Total energies of the molecular cation E^+ and anion E^- (UHF calculations) have been used in evaluating the ionization energy E_i(DSCF) and the electron affinity E_g(DSCF), respectively, according to formulae

$$A \rightarrow A^+ + e^-,\ \Delta E = E(A^+) - E(A) = E_i \quad (1)$$

$$A + e^- \rightarrow A^-,\ \Delta E = E(A^-) - E(A) = E_g \quad (2)$$

In the optimum geometry also the correlation energy has been evaluated *via* the Moller-Plesset 2nd–order perturbative method (MP2): E^0(MP2), E^+(MP2) and E^-(MP2). This allows the evaluation of the corrected ionization energy E_i(MP2) and electron affinity E_g(MP2). In the optimum geometry two molecular graphs have been generated: the molecular electrostatic potential and the electron density; these are plotted at a selected constant contour.

3 Results And Discussion

3.1 Geometry of aminoacids

Aminoacids, in general, can exist or coexist at two basic forms: (i) aminoacid form with the proton attached to the oxygen atom of the carboxyl group; (ii) zwitterionic form with the proton attached to the amine group forming the positively charged ammonium moiety and negatively charged carboxylate site. X-ray structure determination in the solid state confirms the zwitterionic form. In solution, however, these two forms depend upon the pH of the solution; in neutral pH the zwitterionic form is present. The modelling "*in silico*" shows that these two forms are close in energy and the fixed form could depend upon the method of calculation and also the starting structure for the geometry optimization. In general, the PM3 method reproduces the zwitterionic structure of the system. The *ab initio* method often turns the C_a–NH_3^+ group and the carboxylate group –COO^- in the way that a five-membered ring {N-C_a-C-O-H} is formed where the hydrogen atom is attached to the carboxylate oxygen (**Figure 1**). Perhaps polarization functions embodied in the basis set are responsible for such an effect.

The geometries of the X-ray determined molecular structure, optimized geometry, molecular electrostatic potential, and the molecular electron density functions are envisaged in **Appendix 1**. The calculated molecular properties such as ionization energies and electron affinities (at different level of approximation), dipole moment, dipole polarizability volume, molecular surface, and the molecular volume are comprehensively listed in **Table 1**. Some experimental data, such as dissociation constants K_{a1}(carboxylate), K_{a2}(amine), and the octanol/water partition coefficient P are presented in **Table 2**.

Table 1. Results of the *ab initio* calculations [a]

No	Acid	HOMO	LUMO	E_i(DSCF)	E_g(DSCF)	E_i(MP2)	E_g(MP2)				
Abbr.		HOMO	LUMO	Ei	Eg	Eic	Egc	Dip	Pol	Sur	Vol
1	Gly	-250	114	208	114	**238**	**93**	4.56	*30.0*	210	274
2	Ala	-260	109	217	91	**238**	88	5.48	40.1	236	323
3	Asn	-241	96	173	74	205	73	6.68	55.3	273	396
4	Cys	-233	95	207	83	**240**	79	3.79	54.5	269	378
5	Glu	-259	102	202	89	235	86	4.81	61.7	301	442
6	Arg	-216	101	165	90	187	71	7.27	84.8	337	525
7	Phe	-202	90	174	68	207	*64*	4.89	93.0	347	534
8	Tyr	-195	84	167	86	200	71	6.53	96.1	351	546
9	His	-215	101	185	103	216	**94**	**13.0**	75.6	321	475
10	Trp	-174	84	151	61	*132*	84	*2.97*	**117**	389	615

[a] Abbr. Dip – dipole moment m [debye, D], Pol – dipole polarizability a [Å^3], Sur – molecular surface [Å^2], Vol – molecular volume [Å^3]; energy quantities in kcal mol^{-1}. Bold – maximum, Italic – minimum value.

Table 2. Overview of selected experimental data [a]

	Acid	pP	pK_{a1}	pK_{a2}	pK_{a3}
1	Gly	3.21	2.34	9.60	
2	Ala	2.85	2.34	9.69	
3	Asn	3.82	2.02	8.80	
4	Cys	2.49	1.71	8.33	10.8 (SH)
5	Glu	3.69	2.19	9.67	4.15 (OH)
6	Arg	4.20	2.18	9.09	12.1 (NH_2)
7	Phe	1.38	1.83	9.13	
8	Tyr	2.26	2.20	9.11	10.1 (OH)
9	His	3.32	1.78	8.97	
10	Trp	1.06	2.38	9.39	6.04 (NH)

[a] Acidity constants K_{a1} (carboxylic), K_{a2} (amine), K_{a3} (special group), experimental data little vary depending upon source; octanol/water partition coefficient P [12].

3.2 Characteristic properties of individual species

Glycine. The molecule of the glycine crystallizes in the zwitterionic form. However, the geometry optimization by *ab initio* method yields the aminoacid form as the most stable configuration *in vacuo* (**Figure 2**). The ionization energy calculated *via* DSCF approach (208 kcal mol^{-1}) differs substantially from the assumption of the Koopmans theorem according to which $E_i \sim -E$(HOMO) = 250 kcal mol^{-1}. The inclusion of the correlation energy through the 2nd-order perturbation theory gave the corrected value of E_i(MP2) = 238 kcal mol^{-1}. The electron affinities display less discrepancies, however those data are in principle less accurate. The contour diagram drawn on the molecular electrostatic potentials, in the Appendix, shows the acidic (oxygens, red, negative) and basic (hydrogens, blue, positive) sites. The results obtained by the semiempirical PM3 method copy the *ab initio* data, except the LUMO and consequently electron affinities E_g(DSCF) and corrected E_g(MP2). The dipole moment m = 4.6 D, and dipole polarizability volume a = 30 Å^3 adopt expected values for such a slightly polar molecule.

l-alanine. In match with expectations, the molecular properties are very similar to glycine. This molecule is a bit more polar m = 5.5 D, and more polarizable a = 40 Å^3.

l-asparagine. The geometry optimization resulted in the final form showing a hydrogen bond N-H…O that is a part of the five-membered ring {N-C_a-C-O-H}. This form can be classified as “wrapped” or “packed” one. It is even more polar m = 6.7 D, and even more polarizable a = 55 Å^3.

Cysteine. This is a very different molecule whose structure refers to the “open” or “unpacked” aminoacid form. *Ab initio* data show negative value of the LUMO which would indicate a spontaneous reduction. However, the calculated positive electron affinities evaluated *via* eq. (2), E_g(DSCF) = 83 kcal mol^{-1} and E_g(MP2) = 79 kcal mol^{-1}, do not confirm such a predisposition. There is a rather low polarity m = 3.8 D, and medium polarizability a = 54 Å^3.

Glutamic acid. The geometry optimization resulted in the open (unpacked) aminoacid form. The polarity and polarizability are m = 4.8 D, and a = 62 Å^3.

Arginine. While the X-ray structure analysis confirms an unpacked form of this molecule, the geometry optimization resulted in the wrapped zwitterionic form with two hydrogen bonds of the carboxylate oxygen atom to the hydrogen attached to the guanidinium group. Large polarity m = 7.3 D, and enhanced polarizability a = 85 Å^3 are predicted.

l-phenylalanine. This is the first member of the series containing an aromatic ring. The geometry optimization resembles the CCDC pattern, however, with the final packed aminoacid form. Predicted polarity is m = 4.9 D and rather large polarizability a = 93 Å^3.

l-tyrosine. Attachment of the OH group in this molecule does not alter the properties significantly: the geometry converged to the packed aminoacid form with polarity m = 6.5 D and polarizability a = 96 Å^3.

l-histidine. This molecule is the only one that retains its zwitterionic form also *in vacuo*. This causes much increased dipole moment m = 13.0 D but a medium polarizability a = 76 Å^3.

l-tryptophan. The aminoacid form is more stable *in vacuo* than the zwitterionic form. Though the dipole moment is the lowest over the studied series m = 3.0 D, the polarizability is the highest a = 117 Å^3. This molecule displays the lowest ionization energy E_i(MP2) = 132 kcal mol^{-1}.

3.3 Application of multivariate methods

The worksheet formed of data from Tables 1 and 2 was processed by applying the Cluster Analysis (CA), Probabilistic Neural Network Classifier (PNN), Principal Component Analysis (PCA), and the Pearson Correlation (PC).

Results of the CA (Wards method, squared Euclidean norm) are shown in **Figure 3** for the objects and variables. The data show a classification of objects into three clusters **1** = {1, 2, 5} = {Gly, Ala, Glu}, **2** = {3, 6, 9, 4} = {(Asn, Arg, His), Cys}, and **3** = {7, 8, 10} = {Phe, Tyr, Trp}; cysteine could form an own subgroup. A bit surprising is the classification of histidine lying outside the group of other aromatic species. Notice, histidine, arginine and asparagine appear *in vacuo* in the zwitterionic form as opposed to the rest of the studied aminoacids (see Figure 1). They exhibit the largest dipole moment, 13.0 and 7.3, and 6.7 D, respectively.

Concerning the molecular properties, these are grouped according to the similarity into several groups: the group **A** = {LUMO, -HOMO, Ei, Eic} shows a close relationships of variables describing the ionization process; **B** = {Eg, Egc} describe the electron affinity. The pair **C** = {Dip and -logP} refer to the polarity and lipophilicity/hydrophobicity whereas the group **D** = {Pol, Sur, Vol} is associated with the molecular topology. The distinct group **E** = {pKa1, pKa2} refer to the acidity constants that are closely related.

The PNN classifier rearranges the a priori classified input group of the objects (aliphatic and aromatic) into the output group showing the “incorrectly classified” cases (Table 3). Just the object 9 (histidine) has lower “distance” $d = 0.46$ to the aliphatic group and longer $d = 0.54$ to the aromatic group.

Table 3. Overview of selected experimental data [a]

	Acid	Cluster	Input group	Output group
1	Gly	1	al	al
2	Ala	1	al	al
3	Asn	2	al	al
4	Cys	2	al	al
5	Glu	1	al	al
6	Arg	2	al	al
7	Phe	3	ar	ar
8	Tyr	3	ar	ar
9	His	2	**ar**	**al**
10	Trp	3	ar	ar

[a] al – aliphatic, ar – aromatic.

Results of the PCA are displayed in Fig. 4. They give complementary information to the CA. Three variables forming group **D** = {Sur, Pol, Vol} are closely related while the group **A**+**B** = {Eic, Ei, -HOMO, LUMO, Eg, Egc} is positioned at the opposite direction and shows anticorrelation: members of **A**+**B** decrease with increasing **D**. The group **E** = {pKa1, pKa2} is unrelated to **A**+**B**+**D**; **C** = {Dip} is rather singular and anticorrelates with **E**. The objects are spatially distributed into three groups **1** = {1, 2, 5}, **2** = {4, 9, 3, 6} and **3** = {7, 8, 10} in match with the results of CA.

Finally, **Table 4** brings correlation coefficients, *r*, for pairs of molecular properties. These data in a numerical form confirm the results of the PCA. The members of the group **D** show $r \sim 1$, the observable Dip for the group **C** is rather unrelated to the remaining ones.

Table 4. Pair correlation coefficients among molecular properties. [a]

Group		LUMO	Ei	Eg	Eic	Egc	Dip	Pol	Sur	Vol	pKa1	pKa2	pP
A	LUMO												
A	Ei	0.76											
B	Eg	**0.79**	0.60										
A	Eic	0.67	**0.90**	0.64									
B	Egc	0.60	0.53	0.62	0.28								
C	Dip	0.16	-0.08	0.43	0.11	0.21							
D	Pol	**-0.85**	**-0.88**	-0.65	**-0.84**	-0.46	0.02						
D	Sur	**-0.83**	**-0.86**	-0.63	**-0.80**	-0.46	0.08	**0.99**					
D	Vol	**-0.83**	**-0.87**	-0.63	**-0.81**	-0.48	0.07	**0.99**	**1.00**				
E	pKa1	0.19	-0.06	0.09	-0.29	0.26	-0.39	-0.02	-0.07	-0.04			
E	pKa2	0.40	0.19	0.24	0.00	0.41	-0.19	-0.12	-0.13	-0.12	**0.79**		
C	pP	0.64	0.31	0.59	0.44	0.20	0.45	-0.55	-0.47	-0.46	0.01	0.00	
A	mHOMO	**0.83**	**0.86**	0.55	**0.83**	0.40	-0.03	**-0.93**	**-0.88**	**-0.88**	0.07	0.25	0.65

[a] Significant correlation coefficients $r > 0.79$ are bold typed.

4 Conclusions

Starting from the X-ray structure of 10 biogenic aminoacids with the zwitterionic form, the geometry optimization using *ab initio* calculations *in vacuo* results in the amino acid (canonical) form with the hydrogen atom attached not to the amine but to the carboxylate group. There are three exceptions which retain their zwitterionic form: histidine, arginine and asparagine. The statistical multivariate methods confirm a classification of the objects into three groups. Unlike three aromatic aminoacids (Phe, Tyr, Trp), histidine resembles the group of aliphatic aminoacids; it is closely related to Asn and Arg. The classification of the molecular properties is more variable and covers five groups according their similarity: the group **A** = {LUMO, -HOMO, Ei, Eic} shows a close relationships of variables describing the ionization process; **B** = {Eg, Egc} describe the electron affinity. The group **D** = {Pol, Sur, Vol} is associated with the molecular topology the members of the groups **A+B** anticorrelate with those inside the group **D**: with increasing surface, volume and polarizability the ionization energy and electron affinity decreases. The pair **C** = {Dip and -logP} refer to the polarity and lipophilicity/hydrophobicity; the dipole moment correlates only weakly with the partition coefficient -log*P*. The distinct group contains the acidity constants **E** = {pKa1, pKa2} that are mutually closely related and unrelated to the other molecular properties.

Declarations

Supplementary Information The online version contains supplementary material available at (https://doi.org/10.1007/...

Author contributions Both authors contributed to the conceptualization, formal analyses, investigation, writing—original draft, and writing—review and editing the manuscript. R.B. was also responsible for funding acquisition, resources, and supervision.

Funding Slovak grant agencies (APVV 16-0039, VEGA 1/0013/18, ITMS-2014+ No 313011ASN4) are acknowledged for the financial support.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Code availability HyperChem(TM) Professional 7.51, Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

1. V.R. Young, J. Nutr. **1994**, 124, 1517S.
2. R.L. Thurlkill, G.R. Grimsley, J.M. Scholtz, C.N. Pace, Protein Science **2006**, 15, 1214.
3. C.N. Pace, G.R. Grimsley, J.M. Scholtz, J. Biol. Chem. **2009**, 284, 13285.
4. J.-Y. Kim, D.-S. Ahn, S.-W. Park, S. Lee, *RSC Adv*. **2014**, *4*, 16352.
5. A.M. Bayoumy, R. Badry, H.A. Gaber, S.A. Elbiomy, S.G. El Gabaly, M. Sayed Abd ElAziz, S.M. Gouda, H. Elhaes, I.S. Yahia, H.Y. Zahran, M. Ibrahim, *Biointerface Res. Appl. Chem*. **2019**, *9*, 4379.
6. J. Vatrál, R. Boča, W. Linert, *Electrochim. Acta* **2014**, *143*, 53.
7. B. Vranovičová, J. Vatrál, R. Boča, *J. Electroanal. Chem*. **2020**, *860*, 113920.
8. K. Klačanová, P. Fodran, P. Šimon, P. Rapta, R. Boča, V. Jorík, M. Miglierini, E. Kolek, Ľ. Čaplovič, *J. Chem*. **201**3, 961629.
9. L. Kišš, P. Rapta, R. Boča, M. Miglierini, M. Čaplovičová, M. Martinka, L. Žemlička, P. Fodran, *Monatsh. Chem.* **2017**, *148*, 2019.
10. HyperChem – Molecular Modeling System, ver. 8.0.6. Hypercube Inc., 1995 - 2008.
11. Cambridge Crystallographic Data Centre, https://www.ccdc.cam.ac.uk/
12. **PubChem,** National Center for Biotechnology Information, U.S. National Library of Medicine, https://pubchem.ncbi.nlm.nih.gov/compound/Glycine

Figures

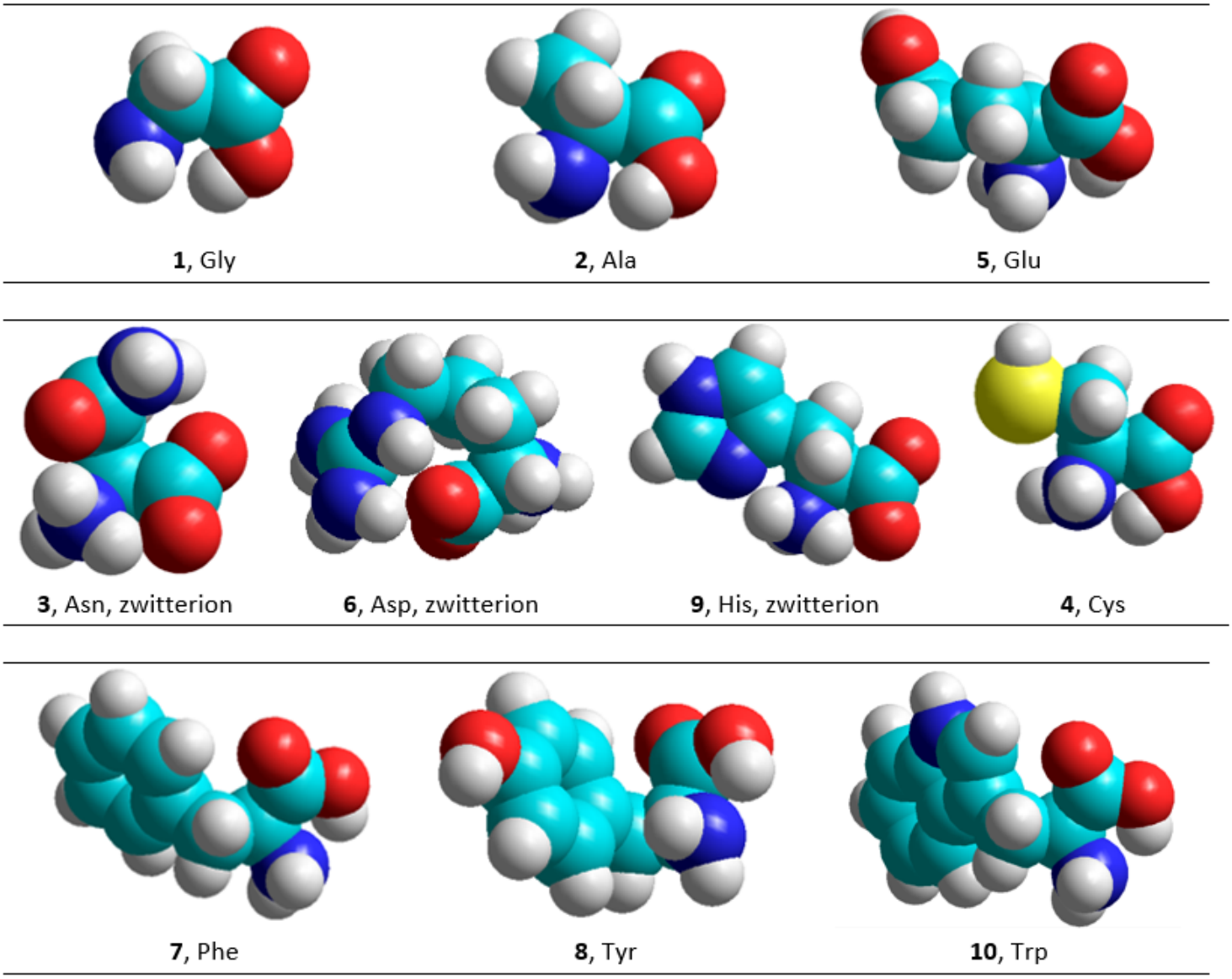
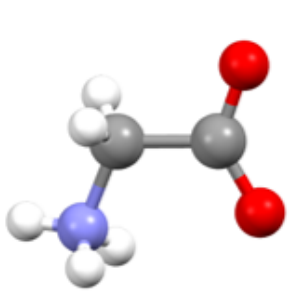


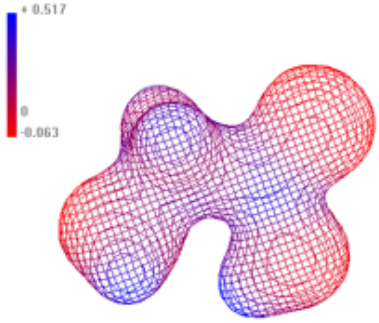
Figure 1

Optimized molecular geometry in vacuo.



CCDC 1515323
zwitterionic form in the solid state

Geometry optimization by *ab initio* method, final aminoacid form *in vacuo*



Molecular electrostatic potential

Figure 2

Geometry and molecular electrostatic potential of the glycine molecule. Analogous information about remaining aminoacids are deposited in Appendix A.

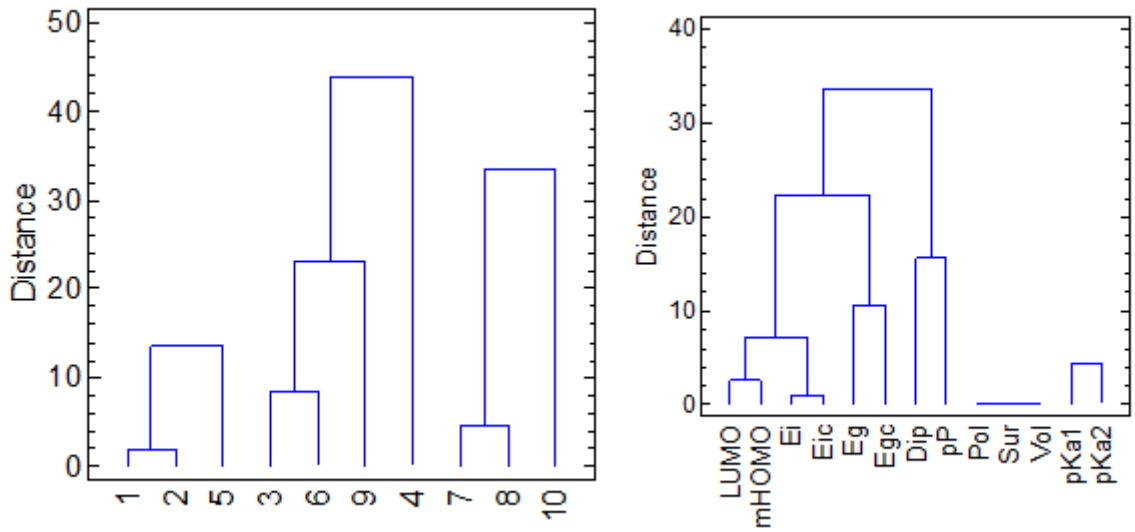


Figure 3

Results of the CA: left – objects (aminoacids), right – variables (molecular properties). mHOMO = -HOMO.

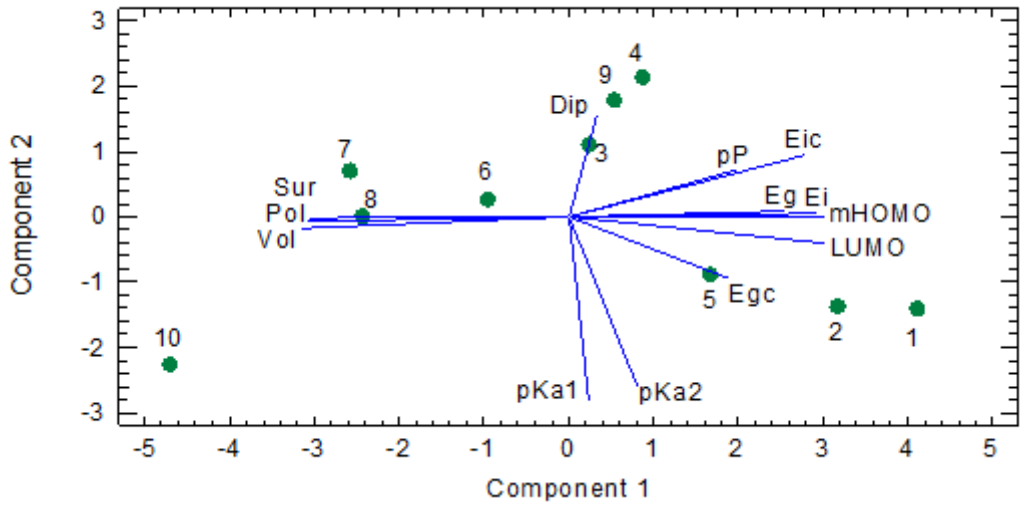


Figure 4

Biplot of principal components by the PCA. The point symbols correspond to the observations, the rays correspond to the variables.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 21ChemaminoESI1701.docx