DISCRIMINANT ANALYSIS OF AMINO ACID COMPOSITION OF NATURAL CARBONACEOUS SUBSTANCES

O. Ye. Amosova, Ye. A. Golubev, S. N. Shanina

Institute of Geology Komi SC UB RAS, Syktyvkar
kramosova@geo.komisc.ru

We conducted statistical estimation of difference in amino acid composition of natural carbonaceous materials, previously divided on the basis of physical and chemical features into groups according to Uspensky classification, by discriminant analysis. The samples of natural solid bitumens from a carbonization sequence (asphaltites — kerites — anthraxolites), natural graphites, and carbon black samples were used as objects of the research. It was shown that the samples within each group had similar amino acid composition and the correctness of an initial distribution of the samples into groups was confirmed. In addition, we found individual samples, which had been incorrectly classified as a result of inaccurate physical-chemical diagnostics, and also slightly different from the samples of their group by the geological and geochemical characteristics. This indicates possibility of using discriminant analysis for the comparative grouping of natural carbonaceous substances by amino acid composition.

Keywords: discriminant analysis, discriminant variables, canonical discriminant functions, natural carbonaceous substances, amino acids.

ДИСКРИМИНАНТНЫЙ АНАЛИЗ АМИНОКИСЛОТНОГО СОСТАВА ПРИРОДНЫХ УГЛЕРОДИСТЫХ ВЕЩЕСТВ

О. Е. Амосова, Е. А. Голубев, С. Н. Шанина

Институт геологии Коми НЦ УрО РАН, Сыктывкар

Методом дискриминантного анализа (ДА) проведена статистическая оценка различия аминокислотного состава природных углеродистых веществ (УВ), предварительно разделенных по физико-химическим признакам на группы в соответствии с классификацией Успенского. Объектами исследований послужили образцы природных твердых битумов из ряда карбонизации (асфальтиты — кериты — антраксолиты), природные графиты, а также образцы технического углерода (сажи). ДА в целом подтвердил правильность первоначального распределения образцов по группам, указав на схожесть аминокислотного состава образцов в каждой из групп. Кроме того, с помощью ДА выявлены единичные образцы, ошибочно отнесённые не к своей группе в результате недостаточно точной физико-химической диагностики, а также незначительно отличающиеся от образцов своей группы по геолого-геохимическим признакам. Это свидетельствует о возможности применения такого анализа для сравнительного группирования природных УВ.

Ключевые слова: дискриминантный анализ, дискриминантные переменные, канонические дискриминантные функции, природные углеродистые вещества, аминокислоты.

Introduction

The method of discriminant analysis of protein amino acid composition is widely used in biological, physiological, palaeontological and geochemical studies. For example, it allows to determine the taxonomic membership of the classes of shells of mollusk, brachiopods and bones of animals [11–13, 15], to identify intergroup differences in biological fluids and pathogenic minerals [1, 2], to determine affiliation to one of the microcommunities in hydrothermal systems [14].

Amino acids, found in sedimentary rocks, are not primary components of protein. They are in the bound state and included in the polymer components of organic matter or they form stable organic mineral complexes [3]. Amino acids are one of the most important indicators of the geochemical processes occurring in sediments. They characterize the degree of polymerization of fossil organic matter that is closely connected with the processes of postsedimentary deposit transformation. Heating effects contribute to both decay of amino acids, especially with a complex structure, and their formation due to the secondary synthesis.

Recently during our studies of amino acids in geological objects we accumulated material on the amino acid composition of different carbonaceous substances from asphaltites [8] to shungites [9]. Previously we made statistical evaluation of the similarity of the samples of shungites from various deposits by their amino acid composition and identification of the impact of shungite formation temperature on their amino acid composition [10].

The relevance of the studies of organic mineraloids is conditioned by debatable questions of their origin, the main of which is the problem of biogenesis or abiogenesis of initial substance, as well as the need of evaluation of physical and chemical conditions of formation.

The complex composition and structure hamper to study of natural solid bitumen and majority of other natural organic compounds. So we continued researches of the amino acid composition of solid bitumen by statistical analysis.

This work aims at the statistical estimation of differences in the amino acid composition of natural carbonaceous substances (CS) that a priori were separated into groups by physical and chemical features according to Uspenskys classification [6].

The objects of the researches were samples of natural solid bitumens from the carbonization sequence (asphaltites (albertite) — kerites (grahamite and impsomite) — anthraxolites), natural graphites, as well as samples of carbon black.

The tasks of the work were as follows:

(1) Establishing the most informative amino acids showing the greatest (statistically significant) differences between the groups of samples;

(2) Identification of statistically significant differences between groups of samples by several most informative amino acids simultaneously;

(3) Creation one or more of canonical discriminant functions (hereinafter - discriminant functions) based on these amino acids that best describe the differences in amino acid composition between CS groups and can be used for including new samples to one group or another.

Objects and methods

We selected 30 samples of natural carbonaceous substances, previously recognized and referred to certain types of organic mineraloids and minerals, as well as two samples of carbon black. The first and second groups included vein (samples 1—3) and clastic (samples 4—12) asphaltites from Timan-Pechora province and Eastern Siberia; the third (samples 13—15) — kerites from Timan-Pechora province; the fourth (samples 16—20) — the higher anthraxolites from Novaya Zemlya and Kazakhstan; the fifth (samples 21—27) — shungites of type I from Karelia; the sixth (samples 28—30) — coal graphite from Taimyr and Karelia, and the seventh (samples 31—32) — carbon black from Sosnogorsk Gas Processing Plant.

The identification and determination of amino acids in the samples were performed with the help of gas chromatograph GC-17A (Shimadzu, capillary column Chirasil-L-Val). The technique is described in [9].

The statistical analysis was carried out by the discriminant analysis (DA) in STATISTICA 6.0 program. The discriminant analysis of CS was conducted by the relative content of amino acids used as discriminant variables. The statistical evaluation was carried out for the level of significance $p = 0.05$.

Results and their discussion

At the first stage we checked the correctness of the initial division of samples into seven groups by thirteen amino acids (alanine (Ala), valine (Val), glycine (Gly), isoleucine (Ile), leucine (Leu), aspartic acid (Asp), glutamic acid (Glu), threonine (Tre), serine (Ser), phenylalanine (Phe), tyrosine (Tyr), proline (Pro), lysine (Lys)). On the basis of classification functions*, which are linear combinations of relative compositions of amino acids, it was determined that all the samples were initially classified correctly.

To identify informative amino acids, making the most significant contribution to the distinction between samples from different groups, we carried out DA by two methods: stepwise with inclusion; stepwise with exclusion [4]. As criteria of informative amino acid selection we used Wilks lambda statistics. The highest percentage of correct classification of samples (97 %) was obtained by stepwise with exclusion. Eight most informative amino acids — Ala, Asp, Tyr, Gly, Leu, Phe, Val, Ile

was revealed by this method. The sets of samples from different groups differ statistically significantly by content of these amino acids. Six discriminant functions were calculated by these amino acids.

The number of discriminant functions, that significantly describe differences between the groups, was determined on the basis of their eigenvalues; relative percentage content of discriminant functions (i.e. percentage of explained dispersion); canonical correlation coefficient; and with the help of the test of significance of residual discriminantability of the system conducted using the Wilks lambda statistics and chi-square test (Table. 1, 2).

The value of eigenvalue and the relative percentage content are connected with discriminating abilities of the function: the larger they are, the greater is the discrimination. All discriminant functions do not necessarily provide a clear discrimination, but their positions are determined in accordance with their individual discriminant abilities. In our case, the first function with eigenvalue 10.6 provides the best discrimination, the second function (4.9) provides a maximal discrimination after the first function, etc. Similarly, if we consider relative percentage contents, it can be seen (Table 1) that the first function contains 51.87 % of the total discriminant abilities, the second — 23.99%, etc. For each subsequent function, this percentage decreases. The last functions are as weak, compared to the first ones, that they do not necessarily introduced a significant contribute to the definition of the differences between the groups.

The published results of DA showed that the feasibility of using discriminant functions for distinguishing objects can be best evaluated by the value of canonical correlation coefficient [4], which is a measure of the bond (the degree of dependence) between the groups and the discriminant function. If this coefficient is small, the discriminant function can be omitted for interpretation.

The coefficient of canonical correlation (Table 1) of the first four discriminant functions has values close to one; it is particularly high for the first two functions, which indicates to a strong dependence between the groups and the first two discriminant functions. The fifth and sixth functions possess relatively weaker relation with the groups. Based on the values of the canonical correlation coefficient, we can say that the groups are clearly different by the studied variables (i.e. relative contents Ala, Val, Gly, Ile, Leu, Asp, Phe, Tyr), otherwise all correlations could have small values. Evaluating relative percentage content and canonical correlations it is possible to determine how many and what discriminant functions are significant in determining differences between the groups.

Another criterion, to determine the number of discriminant functions, needed to determine differences between the groups, is to test the statistical significance of the residual discriminant ability of the system (discriminant functions) before calculating next discriminant function. Eight used discriminant variables are very effectively involved in distinguishing between the groups (Table 2). Wilks lambda close to zero (for $k = 0$) at significance level $p < 0.05$ indicates that there are significant differences between the groups yet before calculating of any of the discriminant functions. Therefore, it can be noted that the results have

Footnote:

* classification function — a linear combination of discriminant variables that maximizes differences between groups, but minimizes dispersion within the groups. For each group its classifying function is determined. Object is classified in the group in which it had the highest value of classification function.

Table 1

The values and significance measures of discriminant functions (DF)

Таблица 1

Собственные значения и меры значимости дискриминантных функций (ДФ)

DF / ДФ	Eigenvalue Собственное значение	Relative percentage Относительное процентное содержание	Canonical correlation Каноническая корреляция
1	10.577	51.87	0.956
2	4.892	23.99	0.911
3	2.295	11.25	0.835
4	1.667	8.17	0.791
5	0.620	3.04	0.619
6	0.341	1.67	0.504

Table 2

Values of Wilks lambda statistics and chi-square test with subsequent excluded discriminant functions

Таблица 2

Значения статистики лямбда Уилкса и критерия Хи-квадрат с последующими исключёнными дискриминантными функциями

k	Wilks' lambda Лямбда Уилкса	Chi-square Хи-квадрат	Degrees of freedom Степени свободы	p-level p-уровень
0	0.0008	168.541	48	0.000
1	0.009	110.990	35	0.000
2	0.052	69.310	24	0.000003
3	0.173	41.289	15	0.0003
4	0.460	18.237	8	0.020
5	0.746	6.901	3	0.075

Note. k — number of subsequently excluded discriminant functions beginning with one.

Примечание. k — количество последовательно исключённых дискриминантных функций, начиная с первой.

been obtained for the samples from the general set with differences between the groups. After excluding five discriminant functions (Table 2) the level of residual discriminant capacity is low (the value of Wilks lambda is close to one, the significance level $p = 0.075$). Consequently, the first five discriminant functions are statistically significant (as a whole as the system of functions). The remaining sixth discriminant function is statistically insignificant, so this function can be omitted.

Using the discriminant functions as a system, we can reduce the information, necessary for the discrimination of the groups, to the smallest number of dimensions, at which the significance of the discrimination is sufficient and the discrimination is correct. In the course of our analysis the dimension of the information space, needed to separate the samples into

groups, decreased from original thirteen discriminant variables to eight most informative variables and further to five discriminant functions.

Now let us determine the relative contribution of individual discriminant variable in the value of each discriminant function. To do this the coefficients should be presented in a standard format (Table 3). The standardized coefficients are used to identify the variables that contribute to the value of the discriminant function. The larger the absolute value of the standardized coefficient of the variable, the greater the contribution of this variable to the value of the discriminant function.

The maximum contribution to the first discriminant function is made by the variables Ala and Gly, the following in significance contribution is introduced by Asp. All other variables are

Table 3

Standardized coefficients of discriminant functions

Таблица 3

Стандартизованные коэффициенты дискриминантных функций

Variable Переменная	DF standardized coefficients / Стандартизованные коэффициенты ДФ					
	DF1 / ДФ1	DF2 / ДФ2	DF3 / ДФ3	DF4 / ДФ4	DF5 / ДФ5	DF6 / ДФ6
Ala	-0.876	0.437	0.091	0.181	-0.221	-0.342
Val	-0.206	0.863	-0.452	0.283	0.329	-0.510
Gly	-0.895	0.344	-0.115	-0.464	-0.080	0.134
Ile	-0.081	1.084	-0.058	0.233	0.157	-0.286
Leu	-0.335	-0.533	-0.680	-0.190	-0.394	-0.520
Asp	-0.567	0.915	-0.544	-0.752	-0.427	0.055
Phe	0.120	0.497	0.749	0.309	-0.813	0.245
Tyr	0.00062	0.118	0.350	-0.735	0.206	-0.627

much less significant. For the second discriminant function three of the eight variables (Ile, Asp, Val) have relatively large standardized coefficients, so they introduce about the same contribution to its value. All other variables are compared to them are secondary. Phe, and Leu are dominant variables for the third discriminant function. But Ala and Ile, which contribute substantially to the first and second discriminant function, on the contrary, do not provide a meaningful contribution to the third discriminant function. Fourth discriminant function is marked by a great contribution of Asp, Tyr, and the fifth — Phe, all other variables are secondary importance.

To determine the closeness of the connection between the individual discriminant variables and discriminant functions, we will consider their correlation (its coefficients are termed complete structural coefficients). At a large absolute value of the coefficient all the information about the discriminant function is contained in the variable.

The first discriminant function (Table 4) is most closely associated with Ala and Gly, the second — with Val, Leu, Asp, the third — with Leu, Asp, Phe, Tyr, the fourth — with Asp, Tyr, the fifth — with Phe, Leu, Ile. Note that the discriminant functions 2, 3 and 4 do not have dominant (large) coefficients. At that the variables that contribute most to the values of the corresponding discriminant functions (Table 3) are generally most closely associated with these functions (Table 4).

Table 5 shows average values of the discriminant functions for each group of samples. Pairwise differences of these values for each function allow determining groups that are the best differentiated by this discriminant function. Figurative points on Figure 1 show geometrical arrangement of the samples rela-

tively to each other at coordinate planes of some pairs of discriminant functions.

Table 5 shows that the discriminant function 1 discriminates mainly kerites and soot (Groups 3 and 7), as its average values for these groups are significantly different from each other and from the average values of other groups. Figure 1a shows that the samples of the seventh group are well differentiated from the samples of other six groups by the first discriminant function. The samples of the third group are different from the samples of other groups, although not as much as the seventh group of samples. The samples of other five groups do not differ so clearly by the first discriminant function. By discriminant function 2 samples of asphaltites (group 2) are significantly different from the samples of other groups, most strongly - from the samples of higher anthraxolites (group 4). The discriminant function 3 most strongly distinguishes kerites from higher anthraxolites (groups 3 and 4). Unlike the discriminant functions 1 and 2, it discriminates asphaltites from shungites (groups 1 and 5) (Fig. 1, b). It can be also used as auxiliary to the discriminant function 1 for separating kerites from shungites (groups 3 and 5). The discriminant function 4 well discriminates coal graphites (group 6) from other groups, especially from asphaltites (group 1), a bit worse from higher anthraxolites (group 4), well from shungites (group 5) (Fig. 1, c). It can be also used as auxiliary to the discriminant function 1 for discrimination of kerites from coal graphite (groups 3 and 6). The discriminant function 5 well discriminates asphaltites from kerites (groups 1 and 3) and can be used as auxiliary to discriminant function 1 for this discrimination. It also discriminates shungites from coal graphites (groups 5 and 6). Therefore, it can be used as auxiliary to discriminant function 4 for distinction of these groups.

Complete structural coefficients

Table 4

Полные структурные коэффициенты

Таблица 4

Variable Переменная	Complete correlations between discriminant variables and discriminant functions Полные корреляции между дискриминантными переменными и дискриминантными функциями					
	DF1 / ДФ1	DF2 / ДФ2	DF3 / ДФ3	DF4 / ДФ4	DF5 / ДФ5	DF6 / ДФ6
Ala	-0.944	-0.006	0.191	0.135	0.017	-0.112
Val	0.442	0.533	-0.175	0.265	-0.111	-0.290
Gly	-0.883	-0.206	0.190	0.013	0.161	0.190
Ile	0.176	0.316	-0.208	0.369	0.408	-0.075
Leu	0.142	-0.559	-0.466	0.185	-0.445	-0.419
Asp	0.375	0.548	-0.418	-0.505	-0.249	0.165
Phe	0.534	0.216	0.411	0.158	-0.636	-0.114
Tyr	0.337	0.030	0.575	-0.576	0.060	-0.466

Average values of discriminant functions for groups

Table 5

Средние значения дискриминантных функций по группам

Таблица 5

Group No. Номер группы	Average values of discriminant functions / Средние значения дискриминантных функций					
	DF1 / ДФ1	DF2 / ДФ2	DF3 / ДФ3	DF4 / ДФ4	DF5 / ДФ5	DF6 / ДФ6
1	0.725	0.098	1.746	-1.634	1.228	-0.854
2	0.404	2.586	-0.792	-0.108	-0.395	-0.167
3	3.169	0.117	2.971	-0.172	-0.704	0.813
4	0.117	-3.075	-1.223	-1.389	-0.541	0.035
5	0.361	-0.318	-0.762	0.769	0.921	0.516
6	0.968	-2.164	0.593	2.559	-0.440	-0.849
7	-10.665	0.086	1.324	0.139	-0.221	0.193

Together five statistically significant discriminant functions well discriminate the seven groups of samples. Using these five discriminant functions allows unambiguously classifying new (previously ungrouped) samples.

In general, the informativeness of discriminant variables and efficiency of discriminant functions to interpret differences between groups are assessed by the percentage of correctly recognized (classified) samples. The classification* can be used with respect to the unknown objects to predict their affiliation to certain groups, and to verify the accuracy of the classification procedures in relation to the known objects [4]. The proportion of correctly classified objects testifies to the accuracy of classification* procedure.

The classification of the samples can be carried out by: classification functions, Mahalanobis distances squares, a posteriori probabilities of affiliation of samples to a group, which are calculated with the help of Mahalanobis distances squares.

The classification procedure uses discriminant variables themselves and/or discriminant functions. The classification procedure resulted in that only one sample (from the first group) was allocated to the other group (Table 6): a posteriori probability 0.49 for shungites (the fifth group), whereas for vein asphaltites (the first group) it is equal to 0.15. Although all other samples were classified by the program in accordance with the original partition into groups, low or not very high values of a posteriori probability of some samples raise doubt in the correctness of the original classification, or point to individual physical-chemical or geological features of these samples. Thus, for sample 12 (clastic asphaltite, Eastern Siberia), a posteriori probability of affiliation to the given group 2 (clastic asphaltites) is 0.55, and to group 5 (shungites) — 0.42. A posteriori probabilities of affiliation to the given group of samples 16 (higher anthraxolite, Novaya Zemlya) from group 4 and sample 23 (shungite I, Nigozero) from group 5 are not high (0.65) compared to those for other samples from these groups.

Then we carried out classification with the help of five statistically significant discriminant functions. The obtained results were more accurate, because the impact of fluctuations was reduced [4]. It was established that all samples were correctly classified (Table 7). The results of the classification and the values of a posteriori probabilities of affiliation to a group, obtained by the discriminant functions and initial discriminant variables, were almost identical: the same samples (1, 12, 16, 23) had low or not very high values of a posteriori probabilities of affiliation to a predetermined group. Although the first sample was referred to the first group, but the probability of affiliation was low — 0.35.

Explanation of such results lies in the geological and geochemical characteristics of these samples. Sample 12 (Eastern Siberia), previously recognized as clastic asphaltite, was selected from a piece of bitumen with a heterogeneous phase content, which was a mixture of asphaltites and lower kerite and possibly higher kerite. This heterogeneity results in ambiguity of physical and chemical diagnostics of the sample and, accordingly, its affiliation to this or the other group of solid bitumen. Sample 16 (anthraxolite from Pavlovskoe deposit, Novaya Zemlya) is different from other anthraxolites samples (group 4) by the lowest degree of metamorphism [5]. If other anthraxo-

Fig. 1. Location of samples in the planes defined by the discriminant function 1 and 2 (a), 1 and 3 (b), 1 and 4 (c). Inset (b) shows the location of the samples of groups 1 and 5. Figure (c) shows the arrangement of samples only for groups 4, 5 and 6. The dashed line separates groups for discriminant function 4

Рис. 1. Расположение образцов на плоскостях, задаваемых дискриминантными функциями 1 и 2 (а), 1 и 3 (б), 1 и 4 (с). На врезке (б) показано расположение образцов групп 1 и 5. На рис. (с) приведено расположение образцов только для групп 4, 5 и 6. Пунктирные линии разделяют группы по дискриминантной функции 4

Footnote:

*Classification — a process that helps the researcher to make a decision: the specified object «belongs to» or «very much like on» the group (class). The decision is made on the basis of information contained in the discriminant variables.

Table 6

Classification of samples by means of a posteriori probability

Таблица 6

Классификация образцов с помощью апостериорных вероятностей

Sample No. Номер образца	A posteriori probabilities / Апостериорные вероятности							
	Group Группа	1 $p^* = 0.094$	2 $p = 0.281$	3 $p = 0.094$	4 $p = 0.156$	5 $p = 0.219$	6 $p = 0.094$	7 $p = 0.063$
1**	1	0.145902	0.342200	0.026250	0.000044	0.485377	0.000227	0.000000
2	1	0.999940	0.000016	0.000000	0.000000	0.000044	0.000000	0.000000
3	1	0.998936	0.000001	0.000701	0.000321	0.000040	0.000001	0.000000
4	2	0.000008	0.995053	0.000000	0.000000	0.004939	0.000000	0.000000
5	2	0.000008	0.997457	0.000000	0.000000	0.002534	0.000001	0.000000
6	2	0.003769	0.880480	0.000205	0.000008	0.115497	0.000040	0.000000
7	2	0.000000	0.999999	0.000000	0.000000	0.000001	0.000000	0.000000
8	2	0.000000	0.999945	0.000006	0.000000	0.000049	0.000000	0.000000
9	2	0.000541	0.999393	0.000000	0.000000	0.000066	0.000000	0.000000
10	2	0.000066	0.956917	0.000008	0.000001	0.043008	0.000000	0.000000
11	2	0.000015	0.972746	0.000000	0.000000	0.027239	0.000000	0.000000
12	2	0.025804	0.552375	0.002468	0.000000	0.419180	0.000172	0.000000
13	3	0.003209	0.000257	0.993735	0.000004	0.002218	0.000577	0.000000
14	3	0.000007	0.000000	0.999993	0.000000	0.000000	0.000000	0.000000
15	3	0.001303	0.000005	0.997811	0.000001	0.000861	0.000020	0.000000
16	4	0.000036	0.000107	0.000000	0.652913	0.346483	0.000461	0.000000
17	4	0.000005	0.000001	0.000000	0.994323	0.005670	0.000001	0.000000
18	4	0.001424	0.000000	0.000001	0.998015	0.000558	0.000002	0.000000
19	4	0.000000	0.000000	0.000000	0.999775	0.000002	0.000223	0.000000
20	4	0.000008	0.000000	0.000000	0.999882	0.000108	0.000002	0.000000
21	5	0.000052	0.000801	0.000002	0.000366	0.992301	0.006479	0.000000
22	5	0.000760	0.003653	0.000010	0.005140	0.990370	0.000067	0.000000
23	5	0.001392	0.347572	0.000000	0.000044	0.650987	0.000005	0.000000
24	5	0.001948	0.007585	0.000007	0.000000	0.988826	0.001633	0.000000
25	5	0.000103	0.066369	0.000000	0.000073	0.931659	0.001796	0.000000
26	5	0.000103	0.000653	0.000000	0.040185	0.955941	0.003117	0.000000
27	5	0.000209	0.000045	0.000001	0.012673	0.980705	0.006367	0.000000
28	6	0.000000	0.000007	0.000001	0.000014	0.037823	0.962154	0.000000
29	6	0.000000	0.000000	0.000000	0.000262	0.000696	0.999042	0.000000
30	6	0.000329	0.000000	0.000141	0.000017	0.003156	0.996356	0.000000
31	7	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
32	7	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000

Notes. *p — a posteriori probabilities of affiliation of samples to groups, proportional to size of groups, ** — incorrect classifications

Примечания. *р — априорные вероятности принадлежности образцов к группам, пропорциональные размерам групп, ** — неправильные классификации.

lites from this group belong to the class of higher anthraxolites, then this sample is closer to the average anthraxolites. Sample 23 (shungite I from Nigozero), in contrast to the primary migration shungites (samples 21, 22, 24—27), refers to redeposited shungites, which organic matter was buried as clots in coastal shallow water zone, and then during tectonic processes transferred as part of tuffaceous detrital materials [7]. This resulted in significantly greater contamination by mineral impurities of this shungite compared to others.

Conclusion

1. The dimension of the information space, needed to discriminate samples into groups, by the multivariate statistical analysis, decreased from original thirteen discriminant variables (amino acids) to eight (the most informative discriminant variables) and then to five discriminant functions, the number

of which was established with the help of such characteristics as canonical correlation coefficient, test of significance of residual discriminant capacity of system.

2. Using standardized coefficients and complete structural coefficients we determined the most significant discriminant variables (no more than three amino acids) for each discriminant function. Herewith the set of discriminant variables is individual for each discriminant function. Each of them allows significantly discriminating two or more groups of natural carbonaceous substances.

3. The results of the classification and values of a posteriori probabilities of group membership, obtained by canonical discriminant functions and initial discriminant variables, are similar. Further statistically significant discriminant functions can be used for a more accurate (compared to discriminant variables) grouping of new (previously ungrouped)

Table 7

Classification of samples by a posteriori probability in the space of two statistically significant discriminant functions

Таблица 7

Классификация образцов по апостериорным вероятностям в пространстве пяти статистически значимых дискриминантных функций

Sample No. Номер образца	A posteriori probabilities / Апостериорные вероятности							
	Group Группа	1 p = 0.094	2 p = 0.281	3 p = 0.094	4 p = 0.156	5 p = 0.219	6 p = 0.094	7 p = 0.063
1	1	0.348381	0.330691	0.015794	0.000035	0.304561	0.000537	0.000000
2	1	0.999455	0.000044	0.000006	0.000000	0.000495	0.000000	0.000000
3	1	0.991720	0.000002	0.007266	0.000789	0.000223	0.000001	0.000000
4	2	0.000007	0.991756	0.000000	0.000000	0.008236	0.000000	0.000000
5	2	0.000007	0.995271	0.000000	0.000000	0.004722	0.000000	0.000000
6	2	0.005058	0.862462	0.000290	0.000008	0.132128	0.000054	0.000000
7	2	0.000000	0.999998	0.000000	0.000000	0.000002	0.000000	0.000000
8	2	0.000000	0.999930	0.000009	0.000000	0.000060	0.000000	0.000000
9	2	0.000529	0.999362	0.000000	0.000000	0.000108	0.000000	0.000000
10	2	0.000269	0.981763	0.000003	0.000001	0.017964	0.000001	0.000000
11	2	0.000013	0.951049	0.000000	0.000000	0.048939	0.000000	0.000000
12	2	0.046447	0.563077	0.002530	0.000000	0.387638	0.000308	0.000000
13	3	0.006054	0.000267	0.990572	0.000004	0.002021	0.001083	0.000000
14	3	0.000006	0.000000	0.999994	0.000000	0.000000	0.000000	0.000000
15	3	0.052178	0.000029	0.945718	0.000005	0.001296	0.000775	0.000000
16	4	0.000140	0.000148	0.000000	0.726777	0.271146	0.001790	0.000000
17	4	0.000016	0.000001	0.000000	0.995736	0.004243	0.000004	0.000000
18	4	0.003027	0.000000	0.000001	0.996453	0.000515	0.000004	0.000000
19	4	0.000000	0.000000	0.000000	0.999955	0.000007	0.000038	0.000000
20	4	0.000014	0.000000	0.000000	0.999874	0.000108	0.000003	0.000000
21	5	0.000090	0.000831	0.000002	0.000357	0.987555	0.011165	0.000000
22	5	0.012203	0.011381	0.000007	0.010817	0.964535	0.001058	0.000000
23	5	0.001755	0.315878	0.000000	0.000040	0.682321	0.000006	0.000000
24	5	0.002125	0.006266	0.000009	0.000000	0.989825	0.001774	0.000000
25	5	0.000109	0.054388	0.000000	0.000060	0.943544	0.001898	0.000000
26	5	0.000285	0.000853	0.000000	0.046006	0.944303	0.008552	0.000000
27	5	0.001306	0.000086	0.000001	0.018939	0.940391	0.039277	0.000000
28	6	0.000000	0.000008	0.000003	0.000019	0.080956	0.919014	0.000000
29	6	0.000000	0.000000	0.000000	0.000218	0.000731	0.999050	0.000000
30	6	0.000321	0.000001	0.001873	0.000049	0.021526	0.976231	0.000000
31	7	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
32	7	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000

samples, because their use reduces influence of random fluctuations.

4. The multivariate statistical analysis of amino acid composition of natural hydrocarbons can find samples, wrongly related to other group as a result of unreliable physical and chemical diagnostics, and also determines samples different from the samples of their group by geological and geochemical characteristics. This indicates the possibility of applying such an analysis for comparative grouping of natural hydrocarbons.

The researches were carried out with the partial support of the Program of Fundamental Researches, UB RAS No. 15-18-5-5, the amino acid composition was determined by analytical equipment of CCU Geoscience Institute of geology KomiSC UB RAS.

References

1. Belskaya L. V., Golovanova O. A. *Matematicheskaya obrabotka rezultatov analiza slyuny cheloveka v usloviyah kam-neobrazovaniya v polosti rta* (Mathematical processing of results

of analysis of human saliva in conditions of lithiasis in oral cavity). *Matematicheskie issledovaniya v estestvennykh naukakh*. Apatity, Geological institute of Kola SC RAS, 24—27 October 2010. Ed. Yu. L. Voitekhovskiy. Apatity, 2010, pp. 158—170.

2. Belskaya L. V., Shukailo E. S., Golovanova O. A. *Primenenie matematicheskikh metodov pri analize rotovoi zhidkosti sportsmenov* (Mathematical methods for analysis of oral fluids of sportsmen). *Matematicheskie issledovaniya v estestvennykh naukakh*. Apatity, Geological institute of Kola SC RAS, 3—6 October 2011. Ed. Yu. L. Voitekhovskiy. Apatity, 2011, pp. 75—78.

3. Drozdova T. V. *Geohimiya aminokislot* (Geochemistry of amino acids). Moscow, Nauka, 1977, 199 pp.

4. Klecka W. R. *Diskriminantnyi analiz* (Discriminant analysis). In: *Faktornyi, diskriminantnyi i klasternyi analiz* (Factorial, discriminant and cluster analysis). Moscow, Finansy i statistika, 1989.

5. Silaev V. I., Ilchenko V. O., Lyutov V. P., Filippov V. N., Golubev E. A., Kovaleva O. V. *Autigennaya psevdomineralizatsiya v antraksolite* (Authigenous pseudomineralization in anthrax-

olite). *Problemy geologii i mineralogii*. Syktyvkar: Geoprint. 2006. pp. 283—314.

6. Uspenskii V. A., Radchenko O. A., Glebovskaya E. A. *Osnovy geneticheskoi klassifikatsii bitumov* (Principles of genetic classification of bitumen). Leningrad, Nedra, 1964, 267 pp.

7. Filippov M. M. *Shungitonosnye porody Onezhskoi struktury* (Shungites from Onezhskaya structure). Petrozavodsk, 2002, 280 pp.

8. Shanina S. N., Bushnev D. A. *Izotopnyi sostav ugleroda aminokislot tverdykh bitumov* (Isotope composition of carbon of amino acids of solid bitumen). 2014, V. 456, No. 5, pp. 586—590.

9. Shanina S. N., Golubev E. A. *Aminokisloty v shungitakh Karelii* (Amino acids in Karelia shungites). Geohimiya, 2010, V. 48, No. 9, pp. 972—987.

10. Shanina S. N., Golubev E. A., Amosova O. E. *Aminokisloty v shungitovom veshchestve dokembriiskikh porod Karelii* (Amino acids in shungite substance of Precambrian rocks of Karelia). In: *Mineralnyi mir: struktura, raznoobrazie, konstitutsiya mineralov, kristallogenezis i mineraloobrazovanie, biomineralnye vzaimodeistviya, evolyutsiya mineraloobrazuyuschiy protsessov* (Mineral world: structure, biomineral interactions, evolution of mineral forming processes). Syktyvkar, Institute of geology Komi SC UB RAS, 2012, pp. 256—270.

11. Arany S., Ohtani S., Yamamoto T., Sugiyama T. Comparison of aspartic acid racemization between mammoth and human dentinal tissues. *Archives of Oral Biology*, 2007. V. 52, pp. 20—25.

12. Endo K., Walton D., Reymont R. A., Curry G. B. Fossil intra-crystalline biomolecules of brachiopod shells: diagenesis and preserved geo-biological information. *Organic Geochemistry*, 1995, V. 23, No. 7, pp. 661—673.

13. Sauheitl L. Direct Amino Acid Uptake by Plants related to Grassland Diversity — methodological and ecological Investigations Dissertation zur Erlangung des Doktorgrades in den Naturwissenschaften (Dr. rer. nat.) an der Fakultät für Biologie / Chemie / Geowissenschaften der Universität Bayreuth Bayreuth, Oktober 2009.

14. Svensson E., Skoog A., Ameng J. P. Concentration and distribution of dissolved amino acids in a shallow hydrothermal system, Vulcano Island (Italy). *Organic Geochemistry*, 2004, V. 35, pp. 1001—1014.

15. Walton D., Cusak M., Curry G. B. Implication of the amino acid composition of recent New Zealand brachiopods. *Palaeontology*, 1993, V. 36, No. 4, pp. 883—896.

Литература

1. Бельская Л. В., Голованова О. А. Математическая обработка результатов анализа слюны человека в условиях камнеобразования в полости рта // Математические исследования в естественных науках: Труды VI Всеросс. (с междунар. участием) научной школы, посвящ. памяти д. ф.-м. н. Р. В. Галиулина. Апатиты, Геологический институт Кольского НЦ РАН. 24—27 октября 2010 г. / Ред. Ю. Л. Войтеховский. Апатиты: Изд-во К & М, 2010. С. 158—170.

2. Бельская Л. В., Шукайло Е. С., Голованова О. А. Применение математических методов при анализе ротовой жидкости спортсменов // Математические исследования в естественных науках: Труды VII Всеросс. (с междунар. участием) научной школы, посвящ. 300-летию со дня рождения М. В. Ломоносова. Апатиты, Геологический институт Кольского НЦ РАН, 3—6 октября 2011 г. / Ред. Ю. Л. Войтеховский. Апатиты: Изд-во К & М, 2011. С. 75—78.

3. Дроздова Т. В. Геохимия аминокислот. М.: Наука, 1977. 199 с.

4. Клекка У. Р. Дискриминантный анализ // Факторный, дискриминантный и кластерный анализ. М.: Финансы и статистика, 1989.

5. Силаев В. И., Ильченко В. О., Лютое В. П., Филиппов В. Н., Голубев Е. А., Ковалева О. В. Аутигенная псевдоминерализация в антраксолите // Проблемы геологии и минералогии. Сыктывкар: Геопринт. 2006. С. 283—314.

6. Успенский В. А., Радченко О. А., Глебовская Е. А. Основы генетической классификации битумов. Л.: Недра, 1964. 267 с. (Тр. ВНИГРИ. Вып. 230).

7. Филиппов М. М. Шунгитоносные породы Онежской структуры. Петрозаводск: Изд-во Карельского НЦ РАН, 2002. 280 с.

8. Шанина С. Н., Бушнев Д. А. Изотопный состав углерода аминокислот твердых битумов // ДАН. 2014. Т. 456, № 5, с. 586—590.

9. Шанина С. Н., Голубев Е. А. Аминокислоты в шунгитах Карелии // Геохимия. 2010. Т. 48. № 9. С. 972—987.

10. Шанина С. Н., Голубев Е. А., Амосова О. Е. Аминокислоты в шунгитовом веществе докембрийских пород Карелии // Минеральный мир: структура, разнообразие, конституция минералов, кристаллогенезис и минералообразование, биоминеральные взаимодействия, эволюция минералообразующих процессов. Сыктывкар: ИГ Коми НЦ УрО РАН, 2012. С. 256—270.

11. Arany S., Ohtani S., Yamamoto T., Sugiyama T. Comparison of aspartic acid racemization between mammoth and human dentinal tissues // *Archives of Oral Biology*, 2007. V. 52, pp. 20—25.

12. Endo K., Walton D., Reymont R. A., Curry G. B. Fossil intra-crystalline biomolecules of brachiopod shells: diagenesis and preserved geo-biological information // *Organic Geochemistry*, 1995. V. 23, № 7. Pp. 661—673.

13. Sauheitl L. Direct Amino Acid Uptake by Plants related to Grassland Diversity — methodological and ecological Investigations Dissertation zur Erlangung des Doktorgrades in den Naturwissenschaften (Dr. rer. nat.) an der Fakultät für Biologie / Chemie / Geowissenschaften der Universität Bayreuth Bayreuth, Oktober 2009.

14. Svensson E., Skoog A., Ameng J. P. Concentration and distribution of dissolved amino acids in a shallow hydrothermal system, Vulcano Island (Italy) // *Organic Geochemistry*, 2004. V. 35. Pp. 1001—1014.

15. Walton D., Cusak M., Curry G. B. Implication of the amino acid composition of recent New Zealand brachiopods // *Palaeontology*, 1993. V. 36, № 4. Pp. 883—896.