Journal of Chemical Science and Engineering
JCSE, 1(1): 1-14
www.scitcentral.com

Original Research: Open Access

Statistical Aspects of the Interrelation between the Biological Activity of Chemical Compounds and their Molecular Structure

Mukhomorov VK*

*Universita degli Studi di Napoli "Federico II" Via Cintia, I-80126, Napoli

Received January 12, 2018; Accepted January 25, 2018; Published September 20, 2018

ABSTRACT

An attempt was made to construct an adequate model of interrelation of radioprotective properties of biologically active chemical compounds with their electronic and information factors. Biological activity (radiation protective effects) of chemical compounds has been analyzed in relation to their electronic sign and the information function. Statistical comparison of qualitative indices has revealed that electronic and information signs the most informative characteristics of the molecules responsible for radiation protective action. Correlation equations are given for electronic and information dependent change in the antiradiation properties of the molecule. Quantitative estimates were made associating the protective efficiency of the chemical compounds under study with variations in the electronic parameters and dose of chemicals.

Keywords: Bioactivity, Statistics, Molecular Structure, Electronic Sign, Information Function, Radioprotector, Statistical Criterion, Contingency, Correlation.

Abbreviation: I.P: Intraperitonel, A.R.P: Antiradiation Protection, RE: Radioprotective Efficiency, RMSE: Root Mean Square Error.

INTRODUCTION

Knowledge of quantitative stochastic interrelation between the chemical structure of a molecules and its biological activity has important theoretical and practical significance. It is necessary both to clarify the mechanism of biochemical action of molecules, and to search for promising new drugs. It is known that the classical apparatus of probability theory and mathematical statistics is the basis of the stochastic simulation of natural phenomena. The main party of such research is to estimation of the closeness of causal relationships between explanatory parameters and response of the biological system.

Causal relationship implies that their recurrence lead to the same consequences. However, a causal relationship can be subject to fluctuations due to random deviations. These fluctuations are due to the uncontrolled and unaccounted factors and are identified by statistical laws.

One of the most relevant issues of modern chemistry of biologically active substances is the problem of creating new effective radioprotectors. The main demands on these drugs are low effective dose, low toxicity and lack of side effects. The existence of side effects significantly limits the practical applicability of radioprotectors. Statistical methods are the most rational in solving problems that are associated with the study of action of a combination of factors on the biosystem. Since the effect of the interaction of drugs with biosystem depends on many conditions, then it has a probabilistic nature. Therefore it is preferable to use a probabilistic model.

It is not always possible to construct an adequate model which describes the relationship of the chemical structure of the compound with its biological activity. If the model is overloaded with a large number of non-essential characteristics use such model becomes almost impossible.

Corresponding author: Mukhomorov VK, Universita degli Studi di Napoli "Federico II" Via Cintia, I-80126, Napoli; Email: vmukhomorov@mail.ru

Citation: Mukhomorov VK.(2018) Statistical Aspects of the Interrelation Between the Biological Activity of Chemical Compounds and their Molecular Structure. J Chem Sci Eng, 1(1): 1-14.

At the same time, nothing can compensate for the shortcomings of the model, if the main link has been lost. Therefore, an adequate model should be as close as possible to simulate the basic properties of chemical compounds. Figuring out of the connection between molecular structure and biological activity will allow carrying out a targeted search for new chemicals, and also can contribute to deciphering the mechanisms of their bioactivity.

METHOD AND DISCUSSION

For a description of the interrelation of bioactivity with molecular structure, we use the descriptors (attributes), the calculation of which requires knowledge of only the structural formula of chemical compounds. We take into account the remark of Alexander P and Bacq Z [1] on the importance of the primary chemical structure of the drug in the mechanism of protection against ionizing radiation.

We use the average number of electrons in the outer shell of atoms as a sign of the molecule [2]:

$$Z = \sum_{i=1}^{N} n_i Z_i / N \qquad (1)$$

where n_i is the number of atoms of *i*-th kind; Z_i is a number of electrons in the outer electron shell. The summation is performed over all the atoms in a molecule *N* is the total number of atoms. In [3] it was shown that the empirical pseudopotential can be represented in the following analytical form

$$V(r) = -Ze^2 / r - f(r) + F(r) \qquad (2)$$

where $f(r)$ and $F(r)$ are amendments to the Coulomb potential. Amendments depend on the distance *r* between the molecule and the electron.

Two groups of chemical compounds are given in **Table 1** [4,5]. The first group contains chemical compounds with an effective radioprotective effect (dose $\leq$ 1mM / kg; the survival of more than 50%, chemical compounds are marked with "+" sign). The second group contains chemical compounds, which have no anti-radiation activity at high doses: Dose> 2mM / kg (these chemicals are marked with "-" sign). This choice of the chemical compounds imposes restriction on the size of the sample.

Our goal is to find a classification rule that statistically reliable divides the active and non-active chemical compounds. To do this, we use the association method (statistical methods for rates and proportions) for signs which have an alternative variation ("yes" or "no"). Observations and sign (*Z*) of molecules can be represented as 2 × 2 table or tetrachoric table (**Table 2**). We will carry out the analysis of the interrelation of chemical compounds bioactivity and the magnitude of sign of *Z*.

First of all, we need to set the threshold value of the sign $Z^{(th)}$ which statistically significant separates effective radioprotectors from ineffective radioprotectors. We first determine the mean value of the sign of *Z* for the sample chemical compounds (**Table 1**). We obtained the following statistics for average value *Z*:

$$N = 100,\ Z^{(av)} = 2.87 \pm 0.08,\ Z^{(min)} = 2.235,\ Z^{(max)} = 4.462,\ S_z = 0.40. \qquad (3)$$

Here $Z^{(min)}$ and $Z^{(max)}$ are the minimum and maximum values of the sign *Z*; S_z is the standard deviation of the sample. The average value of $Z^{(av)}$ should be compatible with other units of the sample. Typically, the maximum and minimum sample units are questionable. The element of set is out-of-order of the set, if the following inequality holds:

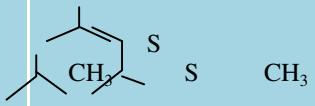
$$\tau = | Z^{(1;2)}_{max/min} - Z^{(av)}_{1;2} | / S_z > \tau^{(cr)}_{1-\alpha}(f) \qquad (4)$$

where f is the number of degrees of freedom. $\tau^{(cr)}_{1-\alpha}(f)$ is the table value of fractile of τ-distribution of the maximum deviation [6]. Let's verify the compatibility of sample points:

$$\tau = | Z^{(max/min)} - Z^{(av)} | / S_z = \begin{cases} 3.99(max) > \tau^{(cr)}_{0.05}(N=100) = 3.40, \\ 1.59(min) < \tau^{(cr)}_{0.05}(N=100) = 3.40. \end{cases} \qquad (5)$$

Table 1: Electronic and information factors of chemical compounds

N	**Chemical compounds**	**I.P. Doze, mM/kg, [4,5]**	**A.R.P.[4,5]**	$Z^{*)}$	***H, bit***
1	$H_2N(CH_2)_4CH(NH)_2CH_2SH$	0.34	+	2.346	1.460
2	$H_2NC(=NH)CH_2CH_2SH$	0.61	+	2.571	1.611
3	$H_2NCH_2CH_2SCN$	0.49	+	2.833	1.730

4	$H_2NCH_2CH_2CH_2NHCH_2CH_2SH$	0.56	+	2.273	1.418
5	$(CH_3)_2NC(=NH)CH_2SH$	0.85	+	2.471	1.545
6	$(CH_3)_3CNHCSNHCH_2CH_2OH$	0.71	+	2.444	1.583
7	$CH_2=CHCH_2NHCH_2CH_2SH$	0.85	+	2.333	1.411
8	$CH_3CH_2CH(NH_2)CH_2SH$	0.95	+	2.235	1.378
9	$(CH_3)_2CH(CH_2)_5NH(CH_2)S_2O_3H$	0.07	+	2.556	1.628
10	$CH_3(CH_2)_6NH(CH_2)_2S_2O_3H$	0.29	+	2.556	1.628
11	$H_2C=C(CH_3)CH_2SC(=NH)NH_2$	0.31	+	2.556	1.568
12	$CH_3NH(CH_2)_3NHCH_2CH_2CH_2SPO_3H_2$	0.31	+	2.606	1.798
13	$H_2N(CH_2)_5NHCH_2CH_2SPO_3H_2$	0.62	+	2.606	1.798
14	$H_2NCH_2C(CH_3)_2CH_2NHCH_2CH_2SPO_3H_2$	0.62	+	2.606	1.798
15	$CH_2=C(NH_2)CH_2CH_2SH$	0.15	+	2.400	1.472
16	$H_2N(CH_2)_5CH(NH_2)CH_2SPO_3H_2$	0.21	+	2.606	1.798
17	Cyclo-$C_6H_{11}NHP(O)(OH)SH$	0.19	+	2.741	1.818
18	$H_2N(CH_2)_3NHCH_2CH_2CH_2SPO_3H_2$	0.32	+	2.667	1.849
19	$H_2NCH_2CH(CH_3)CH_2NHCH_2CH_2SPO_3H_2$	0.44	+	2.667	1.849
20	$H_2NCH_2CH_2CH(CH_3)NHCH_2CH_2SPO_3H_2$	0.66	+	2.667	1.849
21	L(+)=$H_2N(CH_2)_4CH(NH_2)CH_2SPO_3H_2$	0.14	+	2.667	1.849
22	$H_2N(=NH)CH_2SSCH_2CH_2(=NH)NH_2$	0.07	+	2.667	1.641
23	$H_2NCH_2CH_2CH_2NHCH_2CH_2SPO_3H_2$	0.07	+	2.741	1.904
24	$H_2NC(=NH)NHCH_2CH(CH_3)$ $CH_2NH(CH_2)SPO_3H_2$	0.07	+	2.813	1.945
25	$H_2NC(=NH)NH(CH_2)_3NH(CH_2)_3SPO_3H_2$	0.08	+	2.743	1.897
26	$H_2NC(=NH)CH_2SH$	0.13	+	2.727	1.686
27	$H_2N(CH_2)_3NHCH_2CH(OH)CH_2SPO_3H_2$	0.82	+	2.774	1.890
28	$CH_3CH_2CH_2CH_2NHP(O)(OH)SH$	0.15	+	2.667	1.868
29	$H_2C=CHCH_2NHCH_2CH_2SH$	0.85	+	2.333	1.411
30	$H_2NCH_2CH(OH)CH_2NHCH_2CH_2SPO_3H_2$	0.33	+	2.857	1.943
31	$H_2NC(=NH)NHCH_2CH_2NHCH_2CH_2SPO_3H_2$	0.10	+	2.897	1.997
32		0.10	+	2.813	1.649
33		0.05	+	2.813	1.649
34	$H_2N\ CH_2CH_2SCN$	0.49	+	2.833	1.730
35	$H_2NCH_2CH_2SSCH_2CH_2NH_2$	0.99	+	2.500	1.571
36	$H_2N(NH)CNHCH_2CH_2S_2O_3H$	0.50	+	3.300	2.082
37	$H_2N(CH_2)_3NH(CH_2)_3SPO_5H_2$	0.31	+	2.875	1.919
38	$H_2N\text{-}(CH_2)_4NH(CH_2)_2SPO_5H_2$	0.77	+	2.875	1.919
39	$CH_3CONHCH_2CH_2SS(CH_2)_4SO_2H$	0.17	+	2.813	1.781
40	$H_2NCH_2CH_2SS\ CH_2CONH_2$	0.60	+	2.727	1.794
41	L(-)-$H_2NCH_2CH_2CH(NH_2)CH_2SPO_3H_2$	0.63	+	2.833	1.966
42	$H_2NC(=NH)NH(CH_2)_3NHCH_2CH_2SPO_3H$	0.10	+	2.813	1.945
43	$HO_2S(CH_2)_4\text{-}SSS\text{-}(CH_2)_4SO_2H$	0.06	+	2.971	1.739
44	$H_2O_3PSCH_2CH_2NH(CH_2)_3NH\ CH_2CH_2$	0.35	+	2.974	2.014

	SPO_3H_2				
45	$H_2NCH_2CH_2SSCH_2COOH$	0.30	+	3.000	1.918
46	$CH_3S(CH_2)_3NHC(=NH)CH_2S_2O_3H$	0.19	+	3.000	1.939
47	$H_2NCH_2CH(NH_2)CH_2SPO_3H_2$	1.00	+	2.952	2.032
48	$H_2NCH_2CH_2SPO_3H_2$	0.64	+	3.125	2.078
49	$H_2NC(=NH)NHCH_2CH_2SPO_3H_2$	0.25	+	3.143	2.131
50	$HSCH_2CONHNHCOCH_2SH$	0.83	+	3.222	2.059
51	Histamine (H-имидазол-4-этанамин)	0.90	+	2.588	1.447
52	Mexaminum	0.05	+	2.643	1.473
53	Serotonin (5-hydroxytryptamine)	0.06	+	2.720	1.514
54	Thiazolidin	0.85	+	2.333	1.411
55	$H_2NCH_2CH_2CH_2CH_2CH_2SSH$	0.24	+	2.381	1.454
56	$H_2NCH_2CH_2CH_2NHCH_2CH_2CH_2SPO_3H$	0.33	+	2.724	1.883
57	$(CH_3)_2NC_6H_4CH(OH)S(CH_2)NH_2$	0.78	+	2.600	1.600
58	$CH_2=CHCH_2NHCSNH_2$	6.89	-	2.667	1.640
59	$CH_3CH(NH_2)COSH$	11.4	-	2.769	1.823
60	$H_2NCH_2CH_2SO_2NH_2$	4.83	-	2.933	1.907
61	H_2NSSO_3H	4.65	-	4.222	1.891
62	H_2NCH_2COSH	11.0	-	3.000	1.961
63	$CH_3CH_2CH_2NHCSNH_2$	4.23	-	2.471	1.545
64	$HCONHCH_2CH(CH_3)SH$	3.36	-	2.625	1.717
65	$H_2NCH_2CH_2COSH$	9.51	-	2.769	1.823
66	(CH3)C(SH)CH(NH2)COOH	13.4	-	2.938	1.875
67	$(CH_3)_2NCSSH$	4.12	-	2.769	1.669
68	CH_3 $CH(NH_2)COSH$	11.4	-	2.769	1.823
69	$H_2NCOCH(NH_2)CH_2SH$	9.99	-	2.800	1.857
70	$H_2NC(=NH)SCH_2CH_2CH_2SO_3H$	10.1	-	3.143	2.012
71	$(CH_3)_2$ $NNHCH_2CH_2SH$	4.16	-	2.316	1.457
72	$CH_3CH_2OCOCH_2NHCSSCH_2CH_3$	5.07	-	2.522	1.491
73	$H_2C=CHCH_2NHC(O)SCH_2COOCH_2CH_3$	4.93	-	2.846	1.174
74	$HO(CH_2)_2CH_2NHCH_2CH_2S_2O_3H$	3.72	-	2.960	1.855
75	4-(2- Mercaptooxazolyl)-Erythrite	8.97	-	3.000	1.807
76	$H_2NCH_2CH_2SC(O)CH_2$	3.91	-	2.733	1.774
77	$BrC_6H_4O(CH_2)_4NHCH_2CH_2S_2O_3H$	2.13	-	3.000	1.878
78	$CH_3CH_2CH_2NHCSNH_2$	4.23	-	2.471	1.545
79	$CH_3CH_2SC(S)NHCH_2COOH$	5.59	-	3.053	1.925
80	$HO_2CCH_2NHCONHCH_2CH_2SH$	5.62	-	3.048	1.936
81	Tionicotinamide	4.71	-	3.067	1.706
82	$CH_3SC(O)CH_2CH_2NHCONHCH_2CH_2$ $SC(O)SCH_3$	12.3	-	3.031	1.918
83	$HOCH_2(CHOH)_2CH_2NHCH_2CH_2S_2O_3H$	3.07	-	3.067	1.853
84	$HOCH_2CHOHCH_2NHCH_2CH_2S_2O_3H$	7.60	-	3.077	1.880
85	2-Carboxypyrrolidine-1-Dithiocarboxylic acid	5.24	-	3.211	1.958
86	$CH_3OCOCH_2CH_2SO_2CH_2CH(NH_2)COOH$	3.18	-	3.143	1.901
87	$H_2NC(=NH)SCH_2CH_2CH_2SO_3H$	10.1	-	3.143	2.013
88	$[H_2NC(=NH)NHCH(COOH)CH_2S]_2^-$	3.09	-	3.167	2.017
89	N-Oxide 4-Mercaptodihydropyridine	7.87	-	2.970	1.892
90	$H_2NCH_2CHOHCH_2S_2O_3H$	5.35	-	3.263	1.970
91	$H_2NCH_2CH(CH_2OH)S_2O_3H$	4.81	-	3.263	1.970
92	$CH_3C(=NH)SCH_2CH_2CH_2S_2O_3H$	5.08	-	3.130	1.951

93	2-Furyl-$CH_2NHC(=NH)CH_2S_2O_3H$	4.00	-	3.360	2.049
94	$H_2NCONHCH_2CH_2S_2O_3H$	5.00	-	3.474	2.103
95	γ-(S-Purinyl) Thiopropylsulphonic acid	4.42	-	3.407	2.089
96	$HCONHCH_2CH(CH_3)SH$	3.36	-	2.625	1.717
97	$CF_3CF_2CH_2OCOCH_2CH_2NHCH_2CH_2S_2O_3H$	3.00	-	3.818	2.249
98	$(NC)_2C=C(SH)_2$	3.94	-	4.000	1.922
99	1,2,5-Thiadiazole-3-Carboxylic acid	7.69	-	4.146	1.842
100	1,2,5-Thiadiazole -3,4-Dicarboxylic acid	4.60	-	4.462	2.162

*) The number of electrons in the outer shell of an atom: $Z(H) = 1$, $Z(C) = 4$, $Z(N) = 5$, $Z(S) = 6$, $Z(P) = 5$, $Z(O) = 6$, $Z(Pb) = 4$, $Z(Br,F) = 7$.

From the inequality (5) it follows that the chemical compound number $N = 100$ ($Z^{(max)} = 4.462$) is not compatible with other the sample units. Consequently, the chemical compound is to be excluded from the sample and calculating the average value must be repeated. After recurrence the calculations, we have found that the chemical compounds numbered 96, 97, 98, and 99 also must be excluded from the sample. Now the average value has the following statistics:

$$N = 95,\ Z^{(av)} = 2.80 \pm 0.05,\ Z^{(min)} = 2.235,\ Z^{(max)} = 3.474,\ S_z = 0.27, \quad (6)$$

$$\tau = | Z^{(max/min)} - Z^{(av)} | / S_z = \begin{cases} 2.44(max) < \tau_{0.05}^{(cr)} = 3.38, \\ 2.06(min) < \tau_{0.05}^{(cr)} = 3.38. \end{cases}$$

Here $Z^{(min)}$ and $Z^{(max)}$ are the minimum and maximum values of Z in sample that contains $N = 95$ units. Sample satisfies the following inequality:

$$\chi^2(f = 7) = 3.094 < \chi_{0.05}^{2(cr)} = 14.1,\ p = 0.88,\ N = 95. \quad (7)$$

Thus, the sample is uniform and fits the normal distribution. Here p value determines the significance level of criterion which determines the probability of error (~ 10%); f is the number of degrees of freedom. Criterion of Wilk-Shapiro is also satisfied: $W = 0.989 > W_{0,05;95}^{(cr)} = 0.950$.

Now we can determine the average value of Z for the effective and ineffective radioprotectors ($N = 95$). As a result, we obtained the following statistics:

$$N_1 = 57,\ Z_1^{(av)} = 2.71 \pm 0.06,\ Z_1^{(min)} = 2.235,\ Z_1^{(max)} = 3.300,\ S_{z1} = 0.24,$$

$$N_2 = 38,\ Z_2^{(av)} = 2.95 \pm 0.08,\ Z_2^{(min)} = 2.316,\ Z_2^{(max)} = 3.474,\ S_{z2} = 0.27. \quad (8)$$

Values of Z are located around $Z_1^{(av)}$ and $Z_2^{(av)}$ for the effective and ineffective chemical compounds, respectively. Using tabulated values of t - distribution, we can verify whether the distinction in the average values of Z sign ($Z_1^{(av)} > Z_2^{(av)}$) statistically significant. First, we compare the variances of samples: $F = S_{z1}^2 / S_{z2}^2 = 1.34 < F_{0.05}^{(cr)}(f_1 = 56; f_2 = 37) = 1.93$. That is, the distinction of the dispersions is not statistically significant. Then we use the following inequality [7]:

$$| Z_1^{(av)} - Z_2^{(av)} | = 0.242 > t_{0.05}^{(cr)}(f) \left\{ \frac{N[(N_1 - 1)S_{z1}^2 + (N_2 - 1)S_{z2}^2]}{N_1 N_2 (N_1 + N_2 - 2)} \right\}^{1/2} = 0.105,$$

$$f = N_1 + N_2 - 2 = 93. \quad (9)$$

The inequality (9) shows that at the 5% significant level the null hypothesis of equality of average values can be rejected. Consequently, the difference between the average values $Z_1^{(av)}$ and $Z_2^{(av)}$ are statistically significant.

In the first approximation, we can assume that the average value $Z^{(av)} = 2.80$ is a threshold that separates chemical compounds with different radioprotective efficiency. However, it is better to choose the threshold value through repeated testing various Z values close to $Z^{(av)}$ (for example, within the mean error). You can then use the value of Z which results to a more convincing statistical inference. This approach is demonstrated in the search of the

classification rules by statistical methods for rates and proportions.

According to the analysis, it is preferable to choose a threshold is equal to $Z^{(th)} = 2.87$. Importantly, the chemical compounds (*NN* = 97 - 100) have the sign of *Z* noticeably larger than the average value $Z^{(av)}$ and therefore does not violate the inequality: $Z_1^{(av)} > Z_2^{(av)}$.

We need to verify to see whether the separation of chemical compounds into two conditional groups is the result of random factors. Description of classifications, it is convenient to start with the construction of the table of mutual contingency (or association) [8,9] (cross-selection method). **Figure 1** shows the distribution of the chemical compounds by quadrants of the rectangular 2 × 2 table (table of "four fields"). In each cell of the table is indicated the number (frequency) of q_{ij} objects. Obviously, the classification model better describes the phenomenon, the closer the contingency table to diagonal form. In which connection for the objects in each quadrant, we do not assume the existence of a functional mathematical relationship between the dependent variable and the explanatory variable.

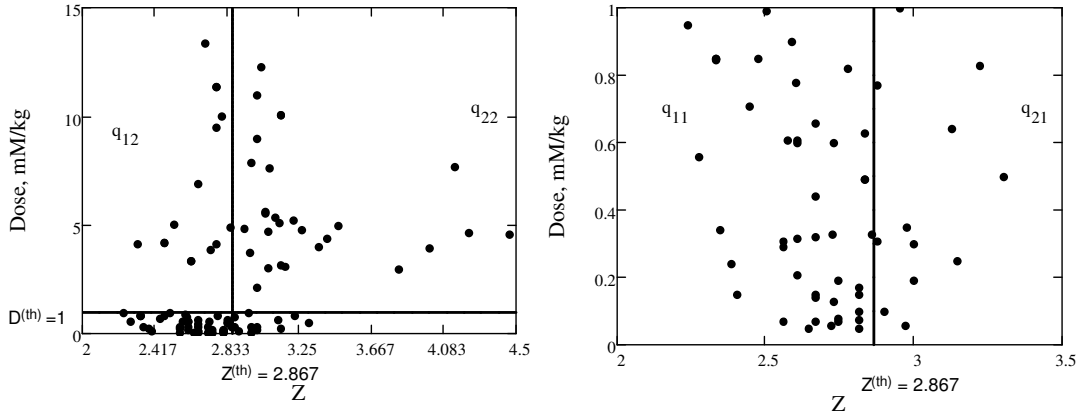


Figure 1. (A). The distribution of chemical compounds in quadrants of table (2 × 2). In the lower left quadrant are chemical compounds (number q_{11}) with signs of $Z \le Z^{(av)}$ and $D \le 1$ mM / kg. In the upper right quadrant are chemical compounds (number q_{22}) with signs of $Z > Z^{(av)}$ and $D > 2$ mM/kg. In the upper left and lower right quadrants specified number of chemical compounds q_{12} and q_{21} with crossed signs: $Z \le Z^{(av)}$, $D > 2$ mM/kg and $Z > Z^{(av)}$, $D \le 1$ mM/kg.
(B). Increased size of the lower two quadrants of the figure (A).

Contingency (association) method is applicable, if the sample size satisfies the following inequality: $N = \sum_{i,j} q_{ij} \ge 40$. It is generally believed that the frequencies q_{ij} meet the inequality of $q_{ij} \ge 5$ subject to $i \ne j$ [8].

We use the following equation to determine the Pearson contingency coefficient Φ [9] between the radioprotective efficacy and value of the sign of *Z* :

$$\Phi = \frac{q_{11}q_{22} - q_{12}q_{21}}{[(q_{11}+q_{12})(q_{21}+q_{22})(q_{11}+q_{21})(q_{12}+q_{22})]^{1/2}} = 0.39 > \Phi_{0.05}^{(cr)}(f = 93) = 0.20 . \quad (10)$$

Here number of degrees of freedom is equal to $f = N - 2$; $q_{11} = 45$ is number of effective chemical compounds having the sign value $Z \le Z^{(th)} = 2.87$ subject to $D \le 1$ mM/kg; $q_{12} = 12$ is number of effective chemical compounds having the sign value $Z > Z^{(th)}$ subject to $D \le 1$ mM/kg; $q_{22} = 29$ is number of effective chemical compounds having the sign value $Z > Z^{(th)}$ subject to $D > 2$mM/kg; $q_{21} = 14$ is number of effective chemical compounds having the sign value $Z \le Z^{(th)} = 2.87$ subject to $D > 2$mM/kg; (**Table 2**). We can also be used the Yule coefficient association for tetrachoric contingency tables [8]:

$$Q = \frac{q_{11}q_{22} - q_{12}q_{21}}{q_{11}q_{22} + q_{12}q_{21}} = 0.77 . \quad (11)$$

The coefficient $Q = 0.77$ point to the existence of the interrelation between the signs. Obviously, this coefficient is in the following range of values: $-1 \le Q \le +1$.

Signs *RE* (the radioprotective efficiency) and *Z* are independent if the product of the marginal or unconditional proportions is equal to the joint proportion (see **Table 2**). For example, we obtained the following result: ${}_1P \times P_2 = 0.57 \times 0.41 = 0.23 \neq p_{12} = 0.13$. These proportions differ considerably. The greater the distinction, the interdependence of signs *RE* and *Z* is greater.

The application of the threshold value $Z^{(th)} = 2.87$ leads to more convincing statistical results than using the average value of In brackets (see **Table 2**), we reported the statistical results that have been obtained for the average value is equal to $Z^{(av)} = 2.80$. Using the average value also suggests a correlation signs at significance level α = 0.05. In this case, the strength of the interrelation too weak: $\Phi = 0.19$. Therefore, it is preferable to use the threshold value 2.87. The adequacy of the model, we can verify using the value of the empirical error. The error is determined by the fraction of misclassified objects: $\Delta = N^{-1} \sum_{i \neq j} q_{ij}$. Using the data in **Table 2**, we found the following value of the empirical error of the model: $\Delta = 0.26(0.33)$. Application of the threshold value $Z^{(th)} = 2.87$ reduces the empirical error of model (approximately 21%).

Table 2. The table of classifications

The sign of *Z*	**Radioprotective efficacy (*RE*)**		**The total sum**
	Effective chemical compound, $D \leq$ 1mM/kg	Inefficient chemical compounds, $D >$ 2mM/kg	
$Z \leq Z^{(th)} = 2.87(2.80)$	$q_{11} = 45$ (36) $p_{11} = 0.45$ (36)	$q_{21} = 14$ (12) $p_{21} = 0.14$ (12)	59 (48) $P_1 = 0.59$ (0.48)
$Z > Z^{(th)} = 2.87(2.80)$	$q_{12} =$ 12 (21) $p_{12} = 0.12$ (0.21)	$q_{22} = 29$ (31) $p_{22} = 0.29$ (0.31)	41 (52) $P_2 = 0.41$ (0.52
The total sum	57 ${}_1P = 0.57$ (0.48)	45(43) ${}_2P = 0.45$ (0.45)	N =100 $\sum_i P =$ $\sum_i P_i =$ 1.00
***Q* = 0.77 (0.63), Φ = 0.39 (0.19), $\chi^2 = 19.9(10.8) > \chi^{2(cr)}_{0.05}(f = 1) = 3.84$ *), *SE* =0.09 (0.09), *K* = 0.43 (0.32), $\lvert r_{tet} \rvert$ = 0.68 (0.53), Δ = 0.26 (0.33).**			

*) Chi-square we calculated using the equation (17).

Let's see the representativeness of the sample (**Table 1**). Using a table of random numbers [6], we will make a partial sample of data **Table 1**. The method of random numbers avoids involuntary and systematic mistakes in the preparation of the sample. As a result, we obtained the following sequence of random numbers:

03, 47, 43, 73, 86, 36, 96, 46, 63, 71, 62, 33, 26, 16, 80, 45, 60, 11, 14, 10, 74, 24, 67, 42, 81, 57, 20, 53, 32, 37, 27, 07, 51, 79, 89, 76, 66, 56, 50, 90. (12)

A series of random numbers, we can obtain, starting from any point of the table of random numbers. We wrote all the random numbers that do not exceed number of 96 [6]. Comparing these numbers with random numbers of chemical compounds **Table 1**, the partial sample of 40 items was obtained. In a partial sample the sequence of chemical compounds represented by "with an open mind" [10]. Statistics of the partial sample is as follows:

$N = 40$, $Z^{(av)} = 2.82 \pm 0.07$, $Z^{(min)} = 2.316$, $Z^{(max)} = 3.300$, $S_z = 0.23$.

$N_1 = 24$, $Z_1^{(av)} = 2.78 \pm 0.08$, $Z_1^{(min)} = 2.333$, $Z_1^{(max)} = 3.300$, $S_{z1} = 0.21$,

$N_2 = 16$, $Z_2^{(av)} = 2.88 \pm 0.13$, $Z_2^{(min)} = 2.316$, $Z_2^{(max)} = 3.263$, $S_{z2} = 0.25$. (13)

This result is similar to the statistics (6), at while the sign of *Z* is represented in the same proportion as in the original sample.

The standard error of contingency coefficient we can be assessed using the following equation:

$$SE(\Phi) = 0.5(1-\Phi^2)(1/q_{11} + 1/q_{12} + 1/q_{21} + 1/q_{22}) = 0.09. \quad (14)$$

Testing of the significance is carried out by using chi - test [9]:

$$\chi^2 = N\Phi^2 = \frac{N(q_{11}q_{22} - q_{12}q_{21})^2}{(q_{11}+q_{12})(q_{21}+q_{22})(q_{11}+q_{21})(q_{12}+q_{22})} = 21.8 > \chi^{2(\mathrm{cr})}_{0.05}(f=1) = 3.84, \quad (15)$$

i.e., at the $\alpha = 0.05$ significance level of the null hypothesis can be rejected. For normally distributed data, you can additionally use the tetrachoric coefficient ($-1 \le r_{\mathrm{tet}} \le 1$) association:

$$r_{\mathrm{tet}} = \cos\left[\pi \cdot (q_{12}q_{21})^{1/2}((q_{11}q_{22})^{1/2} + (q_{12}q_{21})^{1/2})^{-1}\right] = -0.68(-0.53). \quad (16)$$

However, if the distribution of frequencies on borders of two-by-two table is non-uniformly, then coefficient becomes unreliable. Therefore, commonly used [8,9], Pearson goodness of fit (adjusted for continuity of Yates):

$$\chi^2 = \frac{N[(q_{11}q_{22} - q_{12}q_{21}) - N/2]^2}{(q_{11}+q_{12})(q_{21}+q_{22})(q_{11}+q_{21})(q_{12}+q_{22})} = 19.9(10.8) > \chi^{2(\mathrm{cr})}_{0.05}(f=1) = 3.84. \quad (17)$$

Here $N = q_{11} + q_{12} + q_{22} + q_{21}$ is the sum of all frequencies. The inequality (17) shows that there is a statistically significant interrelation of signs. However, the criterion (17) does not give an idea of the strength of the signs interrelation. The assessment of closeness of the linkage between the signs can be obtained by using the coefficient of mutual contingency Pearson:

$$K^2 = \frac{\varphi^2}{1+\varphi^2}. \quad (18)$$

The indicator of mean-square of mutual conjugation φ^2 is equal to:

$$\varphi^2 = \left(\frac{q_{11}^2}{\sum_j q_{1j}\sum_j q_{j1}} + \frac{q_{21}^2}{\sum_j q_{2j}\sum_j q_{j1}} + \frac{q_{12}^2}{\sum_j q_{1j}\sum_j q_{j2}} + \frac{q_{22}^2}{\sum_j q_{2j}\sum_j q_{j2}}\right) - 1 = 0.221. \quad (19)$$

Using equation (18) we determine the coefficient of mutual contingency $K = 0.43$ (0.32), which confirms the interrelation of dichotomous signs.

Study of the interrelationship of the molecules structure - activity showed that the electronic sign of Z is associated with the Shannon informational function [11]:

$$H = -\sum_i p_i \log_2 p_i, \quad (20)$$

where $p_i = n_i / N$, and the following ratios are met for p_i: $0 \le p_i \le 1$, $\sum_i p_i = 1$; $\sum_i n_i = N$, n_i is number of varieties of atoms in the molecule, N is the total number of atoms. The ratio n_i / N determines the relative share of i-th kind of the atom in the molecule [12]. Shannon function is an integral characteristic of the molecule that determines the measure of uncertainty (or diversity) of the structure of chemical compound. The smaller value of the function H, the more diverse (on the relative content of atoms) a multicomponent system.

Using the data of **Table 1** we define the average value of the information function:

$$N = 100, \quad H^{(\mathrm{av})} = 1.80 \pm 0.04, \quad H^{(\mathrm{min})} = 1.174, \quad H^{(\mathrm{max})} = 2.249, \quad S_{\mathrm{H}} = 0.21. \quad (21)$$

We verify the compatibility of the units of the sample on the basis of H:

$$\tau = |H^{\max/\min} - H^{(\mathrm{av})}| / S_{\mathrm{H}} = \begin{cases} 2.15(\max) < \tau^{(\mathrm{cr})}_{0.05}(f=100) = 3.40, \\ 2.97(\min) < \tau^{(\mathrm{cr})}_{0.05}(f=100) = 3.40. \end{cases} \quad (23)$$

Consequently, the sample does not contain incompatible units. Statistics of average values of information functions for effective radioprotectors will be as follows:

$$N_1 = 57, \quad H_1^{(\mathrm{av})} = 1.76 \pm 0.06, \quad H_1^{(\mathrm{min})} = 1.378, \quad H_1^{(\mathrm{max})} = 2.131, \quad S_{\mathrm{H1}} = 0.21. \quad (24)$$

This subset is close to a normal distribution: $\chi^2 = 4.88 < \chi^{2(\mathrm{cr})}_{0.05}(f=4) = 9.49$, and the following inequality satisfies to the criterion of Wilk-Shapiro: $W = 0.951 > W^{(\mathrm{cr})}_{0,05;57} = 0.947$. Let's see the compatibility of the units of this subset:

$$\tau = |H^{\max/\min} - H_1^{(\mathrm{av})}| / S_{\mathrm{H1}} = \begin{cases} 1.78(\max) < \tau^{(\mathrm{cr})}_{0.05}(f=57) = 3.18, \\ 1.80(\min) < \tau^{(\mathrm{cr})}_{0.05}(f=57) = 3.18. \end{cases} \quad (25)$$

These inequalities are point to the lack of incompatible units.

For the inefficient radioprotectors statistics of the average value will be as follows:

$$N_2 = 43, \quad H_2^{(av)} = 1.85 \pm 0.06, \quad H_2^{(min)} = 1.174, \quad H_2^{(max)} = 2.249, \quad S_{H2} = 0.20. \qquad (26)$$

Checking of units of the second subset leads to inequalities:

$$|H^{max/min} - H_2^{(av)}| / S_{H2} = \begin{cases} 2.01(max) < \tau_{0.05}^{(cr)}(f = 43) = 3.10, \\ 3.40(min) > \tau_{0.05}^{(cr)}(f = 43) = 3.10. \end{cases} \qquad (27)$$

From the second inequality (29) it follows that the chemical compound number 16 (H = 1.174 *bit*) is incompatible with the other units of the subset. After excluding this element, we obtained the following statistics for the information function:

$$N_2 = 42, \quad H_2^{(av)} = 1.87 \pm 0.05, \quad H_2^{(min)} = 1.457, \quad H_2^{(max)} = 2.249, \quad S_{H2} = 0.17. \qquad (28)$$

This subset is close to a normal distribution: $\chi^2 = 3.91 < \chi_{0.05}^{2(cr)}(f = 2) = 5.99$. Criterion of Wilk-Shapiro exceeds the critical value: $W = 0.964 > W_{0,05;42}^{(cr)} = 0.942$. The examination of the subset uniformity leads to the following inequalities:

$$|H^{max/min} - H_2^{(av)}| / S_{H2} = \begin{cases} 2.24(max) < \tau_{0.05}^{(cr)}(f = 42) = 3.10, \\ 2.39(min) < \tau_{0.05}^{(cr)}(f = 42) = 3.10. \end{cases} \qquad (29)$$

Thus, the subset comprises only compatible units.

Let's see whether the distinction between the average values of $H_1^{(av)}$ and $H_2^{(av)}$ statistically significant. We predefine the distinction between the variances of S_{H1}^2 and S_{H2}^2: $F = S_{H1}^2 / S_{H2}^2 = 1.52 < F_{0.05}^{(cr)}(f_1 = 56; f_2 = 41) = 1.64$. That is, the distinction in variance is not statistically significant. Therefore, we must use the following inequality:

$$|H_1^{(av)} - H_2^{(av)}| = 0.11 > t_{0.05}^{(cr)}(N-2)\left\{\frac{N_1(N_1-1)S_{H1}^2 + (N_2-1)S_{H2}^2]}{N_1N_2(N_1+N_2-2)}\right\}^{1/2} = 0.07, \qquad (30)$$

$$N = 99, \quad N_1 = 57, \quad N_2 = 42, \quad S_{H1} = 0.21, \quad S_{H2} = 0.17.$$

The inequality (30) rejects the null hypothesis on equality of the average values of the information functions.

Again, we will use the association method of qualitative signs. We choose as the boundary value the following value of the information function (23): $H^{(av)} \equiv H^{(th)} = 1.80bit$. The numerical data are contained in **Table 3**.

Table 3. The table of classifications

The sign of *H*,*bit*	**Radioprotective efficacy (*RE*)**		**The total sum**
	Effective chemical compounds, *D* ≤ 1mM/kg	**Inefficient chemical compounds, *D* > 2mM/kg**	
$H \le H^{(th)} = 1.80$	$q_{11} = 31$ $p_{11} = 0.31$	$q_{21} = 11$ $p_{21} = 0.11$	42 $P_1 = 0.42$
$H > H^{(th)} = 1.80$	$q_{12} = 26$ $p_{12} = 0.26$	$q_{22} = 32$ $p_{22} = 0.32$	58 $P_2 = 0.58$
The total sum	57 $_1P = 0.57$	43 $_2P = 0.43$	$N = 100$ $\sum_i P = \sum_i P_i = 1.00$
$Q = 0.55$, $\Phi = 0.07$, $\chi^2 = 7.20 > \chi_{0.05}^{2(cr)}(f = 1) = 3.84$ [*)], $SE = 0.10$, $K = 0.25$, $\lvert r_{tet}\rvert = 0.46$, $\Delta = 0.37$.			

*) Chi-square we calculated using the equation (17).

Thus, the sign of H serves as the boundary between effective radioprotectors and ineffective chemicals. Variation of the threshold $H^{(th)} \equiv H^{(av)} = 1.80bit$ does not improve the statistical results.

Let's examine these classification rules for chemical compounds that have anti-radiation activity. These chemical compounds were not included in the original sample: 1) $NH_2CH_2CH_2CH_2SH$ (Dose: 3.79mM/kg; Z = 2.73, H = 1.43*bit*), 2) $(CH_3)_2S{=}O$ (Dose: 6.4-12.8mM/kg; Z = 2.60, H = 1.57*bit*), 3) $NH_2CH_2CH_2NHCOCH_2SH$ (Dose: ~ 2mM/kg; Z = 2.63, H = 1.77*bit*), 4) cysteine (Dose: 1.56-1.94mM/kg; Z = 2.36, H = 1.49*bit*), 5) disulfide β – mercaptoethylamine (Dose: 0.99-1.18mM/kg; Z = 2.50, H = 1.57*bit*), 6) S – β aminoethylisothiuronium (AET) (Dose: 1.68-2.10mM/kg; Z = 2.63, H = 1.63*bit*), 7) $(CH_3)_2N$-C_6H_5-CH(OH)-S-$CH_2CH_2NH_2$ (Dose: 0.88-1.77mM/kg; Z = 2.55, H = 1.56*bit*). Obviously, signs of these chemical compounds satisfy the inequalities: $Z < Z^{(th)} = 2.87$, $H < H^{(th)} = 1.80bit$.

The analysis has shown the molecular signs of Z and H are interconnected. For the effective radioprotectors the

interrelation can be described by the following linear regression (**Fig. 2**):

$$H(Z) = A + B \cdot Z, \quad R = 0.87 > R_{0.05}^{(cr)}(f = 55) = 0.22, \quad N_1 = 57, \quad S_1 = 0.122. \qquad (31)$$

The absolute term A and the regression coefficient B are equal to:

$$A = -0.332 \pm 0.338, \quad S_A = 0.169, \quad B = 0.772 \pm 0.124, \quad S_B = 0.062,$$

$$RMSE = 0.109,$$

$$t_B = 12.4 > t_{0.05}^{(cr)}(f = 55) = 2.00 \approx |t_A| = 1.96,$$

$$F = 153.3 >> F_{0.05}^{(cr)}(f_1 = 1; f_2 = 55) = 7.12, \quad t = 9.5 > t_{0.05}^{(cr)}(f = 55) = 1.67. \qquad (32)$$

Here statistics S_1^2 estimates the variance from the regression line; S_A and S_B are standard errors of the regression parameters; R is the sample correlation coefficient. Number of connections is equal to m = 1; number of degrees of freedom is equal to $f = N_1 - m - 1$ [8]. The confidence limits for the free term A and the regression coefficient B at a significance level $\alpha = 0.05$ were determined according to the formula: $t_{0.05}(f = N_1 - m - 1)S_{A,B}$.

For chemical agents which do not possess effective radiation protective action, this interrelation is nonlinear (**Fig. 2**) and can be approximated by the following analytical form:

$$H(Z) = B + A \cdot \exp(-C \cdot Z), \quad A = 3.35 \pm 1.61, \quad B = -(4.06 \pm 0.33), \quad C = 0.3 \pm 0.36,$$

$$N_2 = 43, \quad RMSE = 0.074, \quad F = 85.95 >> F_{0.05}^{(cr)}(f_1 = 1; f_2 = 41) = 4.1. \qquad (33)$$

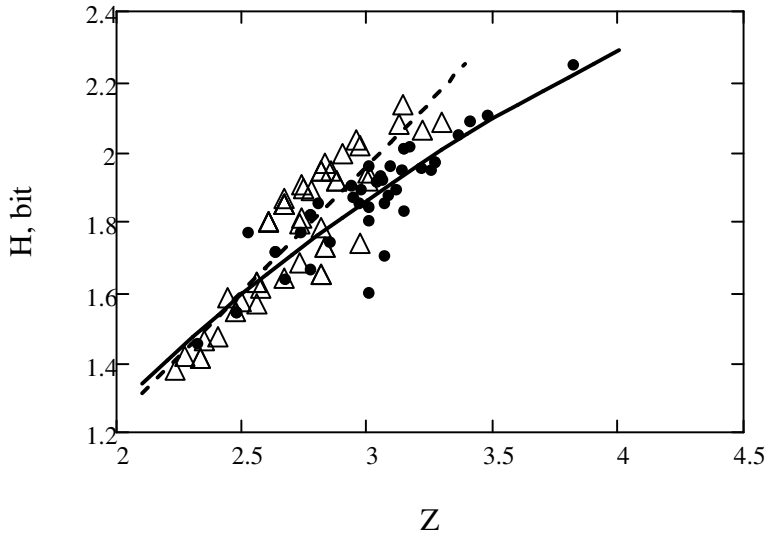


Figure 2. Scattering pattern of the electronic and information signs. H and Z values are taken from **Table. 1**; **- - - -** linear approximation (31). We have marked by triangles (Δ) the effective radioprotectors (N_1 = 57). We have marked by bold dots (•) the chemical compounds that do not have effective radioprotective effect (N_2 = 43). —— non-linear approximation which is defined by the equation (33).

We can get additional information about the nonlinear dependence of $H(Z)$ (Fig. 2) from a variational series of the grouped chemicals. It's typically used 6-8 groups for the sample size $N \approx$ 40-60. You must first make a ranging of the variational series (for example, in ascending of Z). It is convenient to make groups at regular intervals. We chose the number of groups equal to n = 6. Using the approximate relation $\Delta Z \approx (Z^{(\max)} - Z^{(\min)})/n$ we can roughly determine the width of the interval group. Next, we find the group averages $H_{(i)}^{(av)}$ and $Z_{(i)}^{(av)}$ for each classified data. Here i is the group number. Then we compare the ratio of the difference between the average values:

$$b_{HZ,1-2} = \frac{H_{(1)}^{(av)} - H_{(2)}^{(av)}}{Z_{(1)}^{(av)} - Z_{(2)}^{(av)}} = \frac{1.516 - 1.760}{2.419 - 2.733} = 0.78,$$

$$b_{HZ,1-3} = \frac{H_{(1)}^{(av)} - H_{(3)}^{(av)}}{Z_{(1)}^{(av)} - Z_{(3)}^{(av)}} = \frac{1.516 - 1.843}{2.419 - 2.968} = 0.60 \qquad (34)$$

$$b_{HZ,4-5} = \frac{H_{(5)}^{(av)} - H_{(6)}^{(av)}}{Z_{(5)}^{(av)} - Z_{(6)}^{(av)}} = \frac{1.911 - 1.991}{3.100 - 3.269} = 0.47$$

,

$$b_{HZ,5-6} = \frac{H_{(5)}^{(av)} - H_{(6)}^{(av)}}{Z_{(5)}^{(av)} - Z_{(6)}^{(av)}} = \frac{1.911 - 2.119}{3.269 - 3.566} = 0.42$$

.

The subscripts indicate the number of groups. Parameter $b_{HZ,i-j}$ should be close to a constant value for the linear approximation. The frequency of the elements in groups ($3_{(1)}$, $9_{(2)}$, $10_{(3)}$, $13_{(4)}$, $5_{(5)}$, $3_{(6)}$) is close to the normal distribution: $W = 0.902 > W_{0.05}^{(cr)}(n = 6) = 0.788$.

Separation of sample units into groups allows you to calculate the empirical correlation ratio $\eta_{empir.} = \sqrt{S_{HZ}^2 / S_H^2}$ = 0.84. Here S_{HZ}^2 is the between-group variance; S_H^2 is a total variance of the original sample of 43 units. Obviously, the empirical relation $\eta_{empir.}$ changes from zero to one and allows us to quantify the effect of Z factor on the variation of resulting character of *H*.

Then we can calculate the theoretical correlation ratio $\eta_{theor.} = (\delta^2 / S^2)^{0.5}$. Here the value of δ^2 = 0.02 is the variance of equalized values of the information function; S^2 = 0.025 is the variance of empirical (the facts) values of the information function. The theoretical correlation ratio is equal to $\eta_{theor.}$ = 0.89 (coefficient of determination is equal to $\eta_{theor.}^2$ = 0.79). That is, the change of the factor Z (79%) leads to the change of the information function. The remaining change in the information function H (21%) is explained by other factors that were not considered in the model. The nonlinear interrelation between the signs is strong if the inequality 0.7<$\eta_{theor.}$ < 0.9 (scale of Chaddock) is met.

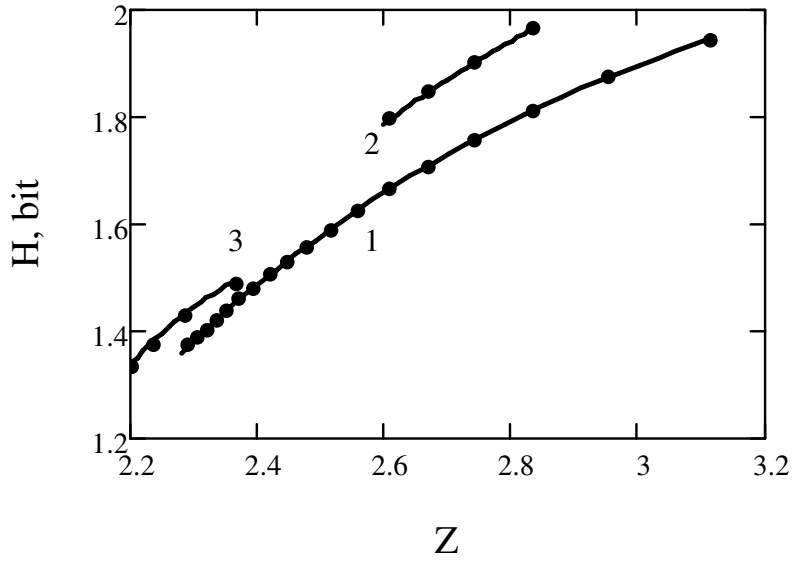


Figure 3. (1) Scattering pattern of the electronic and information signs for radioprotector series: $CH_3(CH_2)_mNHCH_2CH_2SSO_3H$ (m = 0,1,…,17) [4]. The regression equation: $H(Z) = B + A \cdot \exp(-C \cdot Z)$, $A = -15.9 \pm 0.22$, $B = 2.26 \pm 0.003$, $C = 1.26 \pm 0.008$, $RMSE$ = 0.0006. (2) Series of chemical compounds: $CH_3(CH_2)_mNH(CH_2)_nSPO_3H_2$ (m = 2, 3, 4, n = 2, 3) [13]. The regression equation: $H(Z) = B + A \cdot \exp(-C \cdot Z)$, $A = -24.10 \pm 9.15$, $B = 2.42 \pm 0.07$, $C = 1.40 \pm 0.19$, $RMSE$ = 0.0009. (3) Series of chemical compounds: $NH_2(CH_2)_mSH$ (m = 2, 3, 4, 5) [5]. The regression equation: $H(Z) = B + A \cdot \exp(-C \cdot Z)$, $A = -28.4 \pm 17.6$, $B = 1.96 \pm 0.11$, $C = 1.74 \pm 0.36$, $RMSE$ = 0.0011.

Figure 3 shows the interrelationship of the information function and the electronic factor when changing the number of atomic groups CH_2 in molecules. RMSE values are so small, that the interrelation between factors *Z* and *H* come close to a functional interrelation.

As analysis has shown the information function relates to the value of π. The value of π = 0.52 [14] defines an additional contribution of the group atoms CH_2 in hydrophobicity of molecules. **Figure 4** shows this relationship for radioprotectors: $CH_3(CH_2)_mNHCH_2CH_2SSO_3H$ (m = 0, 1, …, 17), $CH_3(CH_2)_mNH(CH_2)_nSPO_3H_2$ (m = 2, 3, 4; n = 2, 3), $NH_2(CH_2)_mSH$ (m = 2, 3, 4, 5).

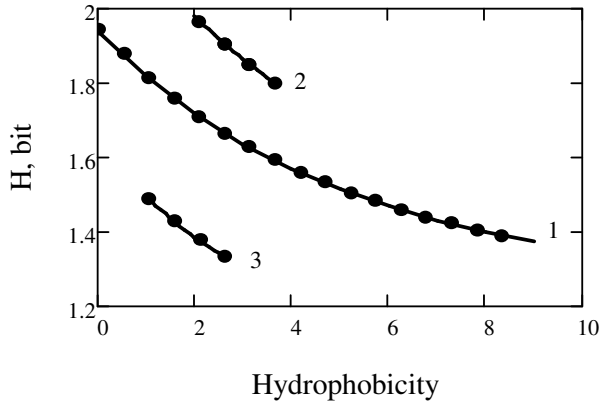


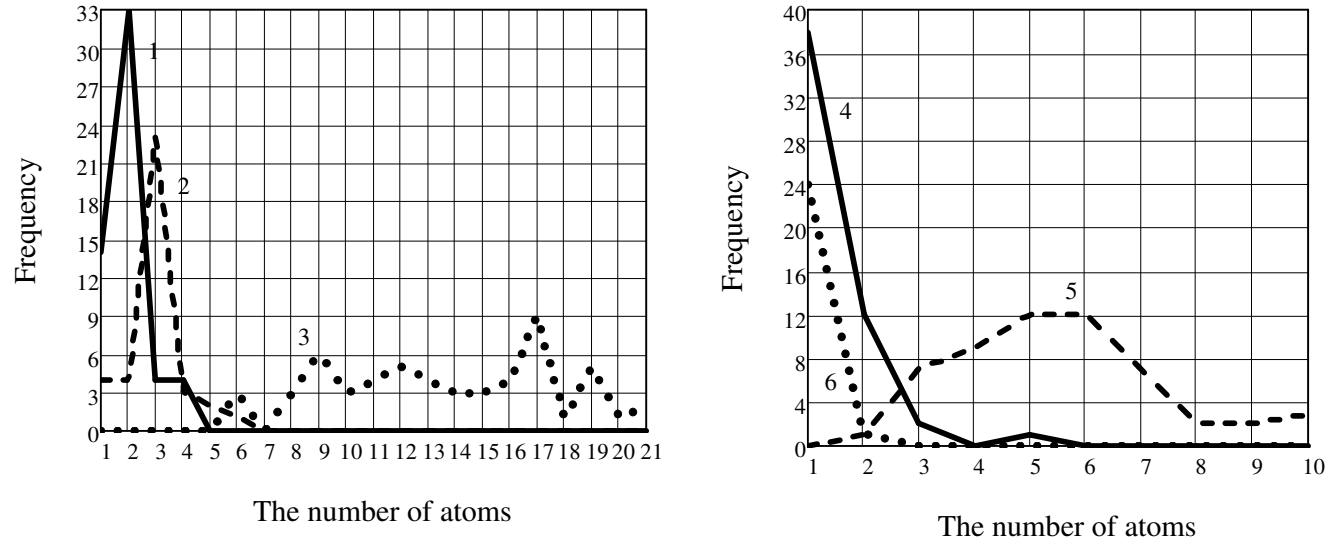
Figure 4. Relationship of the information function with an additional contribution of π in the hydrophobicity of radioprotectors: (1) $CH_3(CH_2)_mNHCH_2CH_2SSO_3H$, (2) $CH_3(CH_2)_mNH(CH_2)_nSPO_3H_2$ and (3) $NH_2(CH_2)_mSH$. (1) The regression line is approximated by the following function: $H(\pi) = A \cdot \exp(-C \cdot \pi) + B$, $A = 0.682 \pm 0.004$, $B = 1.26 \pm 0.004$, $C = 0.198 \pm 0.003$, $RMSE = 0.002$. (2) The regression line is approximated by the following function: $H(\pi) = A \cdot \exp(-C \cdot \pi) + B$, $A = 0.968 \pm 0.030$, $B = 1.32 \pm 0.04$, $C = 0.191 \pm 0.013$, $RMSE = 0.0004$. (3) The regression line is approximated by the following function: $H(\pi) = A \cdot \exp(-C \cdot \pi) + B$, $A = 0.541 \pm 0.015$, $B = 1.11 \pm 0.02$, $C = 0.337 \pm 0.023$, $RMSE = 0.0007$.

The positive interrelation between the signs of Z and H is not random. Information function determines the diversity of the molecular structure, which in turn is determined by the number of different atoms, forming a bound complex of atoms, i.e., molecules. At the same time, the structure of the molecule is not an arbitrary set of various atoms, but is determined by the valence electrons in the outer electron shell. Apparently, this quantum-chemical property establishes the interrelation of two signs of Z and H for molecular structures.

Some distinctions between effective and inefficient radioprotectors we can get if we will analyze the frequency of the atoms appearnce in the molecule. **Figure 5** shows the frequency of occurrence of atoms (C, H, N, O, S, P) in the molecule.

Effective radioprotectors

A B



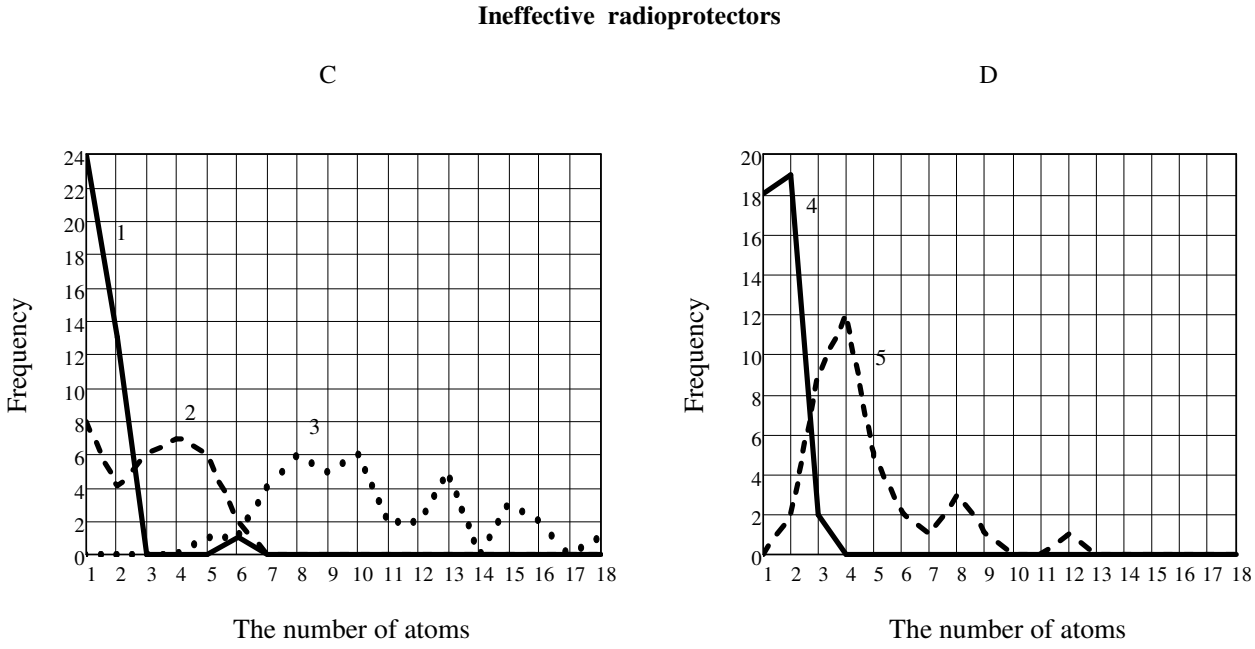


Figure 5. The frequency of appearance nitrogen (1), oxygen (2), hydrogen (3), sulfur (4), carbon (5) and phosphorus (6) in molecules (**Table 1**). In the figure (D) there is no line for phosphorus. In **Table 1** there is only one ineffective agent that contains phosphorus atom.

Using the data of **Table 1**, we can approximately indicate the frequency of occurrence of atoms in a molecule of hypothetical effective agent (for a homogeneous sample): *P* ~ 1, *S* ~ 1, *N* ~ 2, *O* ~ 3, *C* ~ 5-6, *H* ~ 17 (**Fig.5**)[1]. At the same time the most probable distribution of atoms in the inefficient agents (hypothetical molecule) will be as follows: *P* ~ 1, *N* ~ 1, *O* ~ 1, *S* ~ 2, *C* ~ 4, *H* ~ 8-10.

CONCLUSION

The proposed classification rules allow to identify the similarities between the molecular structures. These rules can be practically useful in a preliminary forecast of bioactivity of new chemical compounds. It should be noted that for the calculation of signs of *Z* and *H* is only required the knowledge of the chemical structural formula. This makes much easier the work in a preliminary searching for new bioactive chemicals. Classification rules allow you to set whether you can expect from a chemical compound effective biological action. The ability to separate the biologically active chemical compounds from the inactive chemical compounds on the basis of the sign of *Z*, apparently is due to the existence of the real molecular electrostatic potential. The magnitude of this potential varies from molecule to molecule. Moreover, there is a threshold of the electrostatic potential for effective chemical compounds which is lower of some value (in absolute value). The method described in this article, has yielded positive results when researching antifungal activity and toxicity of chemical compounds [15]. This method was also used in the analysis of the activity of carcinogenic chemicals [16].

However, it should be noted that these rules are not sensitive to iso-electronic molecular systems, as well as for the isomer molecules. This approach gives the most reliable results when analyzing the homologous series of chemical compounds. Homologous series are generally characterized by the signs that satisfy the compatibility condition.

[1] This sequence of numbers is close to the Fibonacci series: 1, 1, 2, 5, 8, 13.

REFERENCES

1. Alexander P, Bacq ZM, Cousens SF, Fox M, Herve A, Lazar J, et al. (1955) Mode of action of some substances which protect against the lethal effects of x-rays. Radiat Res 2: 392.
2. Veljkovič V, Lalovič D (1977) Simple theoretical criterion of chemical carcinogenicity. Experientia 33: 1228.
3. Veljkovič V, Lalovič D (1973) General model pseudopotential for positive ions. Phys Lett A 45: 59.
4. Sweeney TR (1979) A Survey of Compounds from the Antiradiation Drug Development Program. Washington.

5. Romantcev EF (1968) Radiation and chemical protection. Moscow.
6. Handbook of Applicable Mathematics (1984) Vol.VI. Statistics. Part B. John Willey & Sons. Chichester-New York-Brisborne-Toronto-Singapore.
7. Pustyl'nik EI (1978) Statistical methods for the analysis and processing of observations. Moscow.
8. Förster E, Rönz B (1979) Methoden der Korrelations – und Regressionanalyse. Berlin.
9. Fleiss JL (1981) Statistical Methods for Rates and Proportions. Chichester-New York-Brisborne-Toronto-Singapore.
10. Urbach VY (1975) Statistical analysis in biological and medical studies. Moscow.
11. Shannon C (1948) A mathematical theory of communication. Bell Techn J 27: 379.
12. Mukhomorov VK (2012) Modeling of chemical compounds bioactivity. Relationships of structure - bioactivity. Lambert Academic Publisher, Germany.
13. Yaschunsky VG (1975) Progress in the search for chemical protective agents against radiation. Russ Chem Rev 44: 260.
14. Leo A, Hansch C, Elkins D (1971) Partition coefficients and their uses. Chem Rev 71: 525.
15. Mukhomorov VK (2014) Bioactivity-structure. Interrelation of electronic and information factors of biologically activity of chemical compounds. Trends J Sci Res 1: 38.
16. Mukhomorov VK (2011) Entropy approach to the study of biological activity of chemical compounds: The other side of radioprotectors. Adv Biol Chem 1: 1.