Modeling of Radioprotective Activity of Substituted Triptamines

VLADIMIR MUKHOMOROV
Physical Department
State Polytechnic University
St. Petersburg
RUSSIA

Abstract: - A close relationship was established between the bioactivity of substituted tryptaminoes and electronic, geometric, and informational factors. It is shown that for bioactive and inactive chemical compounds there are different interrelationships of electronic information factors. That is, a structural shift was found in the relationships of molecular parameters for bioactive molecules and weakly active drugs.The presence of a structural break does not contradict statistical criteria. The optimal sizes of the substituents have been determined.

Key-Words: - Tryptamine, bioactivity, regression, electronic factor, information factor, structural shift, statistical criteria

Received: May 18, 2021. Revised: March 11, 2022. Accepted: April 13, 2022. Published: May 6, 2022.

1 Introduction

Currently, the two most studied classes of *radio protective* agents include sulfur-containing substances and chemical compounds of a number of indolylalkylamines. Here we analyze the relationship between the radioprotective effect of indolylalkylamines and changes in their molecular structure introduced into the molecule by various substituents. As is known [1,2], substituents have a significant effect on the bioactivity of indolylalkylamines, enhancing or weakening it. Apparently, this is due to a change in the physicochemical properties of molecules and, in particular, changes in their electronic and steric characteristics. Drugs of this class of compounds are classified as radioprotectants of receptor action, the mechanism of which is associated with a hypoxic effect. In this case, the drug leads to a decrease in the delivery of oxygen to the cells and a decrease in its tension in the cytoplasm. Indolylalkylamines are characterized by high psychotropic activity, in addition, they are able to increase the permeability of the blood-brain barrier.

2 Problem Formulation

Z. Bacq [1] indicates that many authors have tried to find for a number of indolylalkylamines the relationship between their chemical structure and radioprotective bioactivity [3-6]. However, the results obtained in these works on the relationship between the variability of the radioprotective efficacy of drugs and changes in the molecular structure cannot be considered satisfactory.

It is known that, in terms of their pharmacological action, these compounds belong to vasoconstrictor drugs that cause hypoxia in the body. In this case, the drug leads to a decrease in the delivery of oxygen to the cells and a decrease in its tension in the cytoplasm. Revealing the relationship between the structure of a molecule and its antiradiation activity would allow not only a targeted search for new drugs, but such studies are also important for a more detailed deciphering of the mechanism of their antiradiation action. Perhaps the results of such studies will create preconditions for the creation of new effective drugs in the series of indolylalkylamines.

3 Problem Solution

The electronic structure of the ground state of indolylalkylamines was analyzed by the semiempirical self-consistent Hartree-Fock method in the MINDO/3 approximation (*Modified Intermediate Neglect of Differential Overlap*), taking into account the optimization of the spatial geometry of molecules [7]. The method gives satisfactory results for most of the standard electronic characteristics of molecules. The excited electronic states of molecules were determined by the method of superimposing electronic configurations CNDO/S (*Complete Neglect of Differential Overlap*). It uses the methods of applied

statistical analysis to compare various electronic characteristics of chemical compounds with the protective efficacy of drugs.

Here we will use the experimental data [2] on the antiradiation protection of substituted indolylalkylamines. Indolylalkylamines are derivatives of 3-(aminoalkyl) indole (Fig. 1). The drugs were chosen in such a way that the model took into account as many different substituents as possible. To maintain the homogeneity of the experimental material, the samples mostly contain drugs that were used in doses equimolar to 50 mg / kg tryptamine. The experiment [2] was carried out on mice with intraperitoneal administration of drugs. Various possible mechanisms of the protective action of substituted tryptamines have been repeatedly discussed in the literature [1,2]. The presence of a substituent in the fifth position of the benzene ring makes these compounds similar to serotonin (5-hydroxytryptamine). It is known serotonin interacts with specific receptors in the body and is highly reactive.

The statistical approach used here to describe the chemical structure - biological activity relationship is based on the following premises. 1) The property of chemical compounds to have a pharmacological effect is determined by the ability to complementarily fix and interact with a certain receptor. 2) The effectiveness of the interaction of a chemical compound with a receptor is determined by the properties of some local regions of the molecule.

Binding of a drug to a receptor is both recognition and initiation of a chain of phenomena leading to a biochemical and physiological response and, ultimately, to a change in the body's resistance. This interaction is limited to the formation of relatively weak quasi-chemical bonds. Thus, it is possible that the biological activity of the drug depends on its affinity for the active center, which, in turn, can be determined by the electronic and steric (geometric) properties of the molecule. The use of a statistical approach to modeling the bioactivity of indolylalkylamines is due to the fact that the drug - organism complex must be considered as a complex system in which many not always controlled processes take place.

The chemical compounds shown in Table 1 can be divided into two groups. The first group of chemical compounds: Nos. 1–25. In this case, the substituents are attached to the indole ring, and the side chain remains unchanged. Changes in the side chain are taken into account in the second group of chemical compounds Nos. 26–35 (Table 3). The integral effect (survival rate A,%) of the body's reaction characterizes the radioprotective activity of a chemical compound.

From a comparative analysis of various electronic parameters and their combinations, it follows that for the first group of compounds, variations in the protective effect of drugs are due to a change, mainly, of two electronic factors of the molecule. First, by changing the value of the charge q_5 on the carbon atom (local factor) in the fifth position of the benzene ring. Second, a change in the position on the energy scale of the level of the lower free molecular orbital $\varepsilon_{\text{unoc}}$ (in units of eV). Third, by variations in the spatial size of the substituents (local factor). For comparability of the results of the analysis, the statistical method suggests [8] to preliminarily perform the division of the initial data into groupings, allowing to exclude any one of the assumed explanatory variables. The distribution of the sets of observational data and the estimated electronic parameters of the molecules presented in Table 1 is close to the normal distribution.

Calculations have shown that substituents in the fourth, sixth and seventh positions of the benzene ring reduce (in absolute value) the orbital electronic charge $q_5 = 2e\Sigma C_{5i}^2$ on the carbon atom of the indole ring (summation is performed over all occupied molecular orbitals (MO)); C_{5i} are the coefficients of the MO decomposition into atomic orbitals. Quantum-mechanical calculations of the electronic structure of molecules showed that the electronic charge is distributed over the carbon atoms of the benzene ring in such a way that it seems to oscillate depending on the position of the substituent in the benzene ring and the donor-acceptor properties of the substituent.

Halogens are strong electron acceptors (the total effect of the inductive and mesomeric shift of the electron density is negative). Electron acceptors usually include substituents COOH, CHO, NO_2, SO_3H. Electron donors include, for example, substituents NH_2, OH, CH_3, C_2H_5 (the total inductive and mesomeric effect is positive). The inductive effect determines the shift of the electron density along the σ bonds and noticeably weakens with increasing distance from the substituent. The mesomeric effect will be positive if the electron density is shifted from the substituent to the system of conjugated π bonds. As shown by the analysis of various electronic parameters and their combinations for tryptamine derivatives, the most informative factor was the combined electronic trait $Q_5 = sign(q_5)(\varepsilon_{\text{unoc}}{}^2 \cdot |q_5|^{0.5}$ (in arb. units) [9, 10]. Here $\varepsilon_{\text{unoc}}$ is the energy of the lowest unoccupied MO (in units of eV). The set of mutually independent variables Q_5 is shown in Table 1.

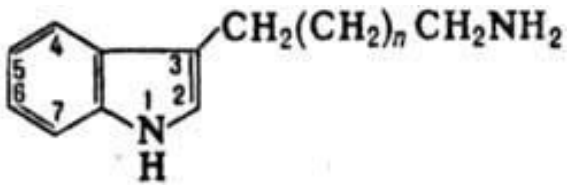

Fig. 1. Chemical structure of 3- (aminoalkyl) indole molecule. Tryptamine corresponds to n = 0 (molecular formula $C_{10}H_{12}N_2$).

Table 1
Radioprotective activity of mice [2] irradiated at a dose of 700 R. Electronic and steric parameters of substituted tryptamines

N	Chemical compounds	$-\varepsilon_{hoc}$, eV	ε_{unoc}, eV	q_5	Q_5, arb.units	*MR*(R_5), cm^3/mol	*Z*, arb. units	*H*, *bits*	Dose, mg/kg	Act_{exp}, %
1	5- Fluorotryptamine R_5 = F	7.998	-0.865	-0.3389	-0.5036	0.92	2.833	1.532	56	67.5
2	5- Methoxy-7- chlorotryptamine $R_5 = OCH_3$, R_7 = Cl	7.882	-1.092	-0.1986	-0.4866	7.87	2.857	1.659	70	64.4
3	5- Chlorotryptamine R_5 = Cl	8.017	-0.966	-0.2240	-0.4572	6.03	2.833	1.532	60	68.0
4	5- Iodtryptamine R_5 = J	8.200	-1.011	-0.1972	-0.4441	13.94	2.833	1.532	89	59.9
5	5- Bromtryptamine R_5 = Br	8.100	-0.952	-0.2011	-0.4269	8.88	2.833	1.532	74	58.0
6	5- Methoxytryptamine (mexamine) $R_5 = OCH_3$	7.768	-0.600	-0.2121	-0.2767	7.87	2.556	1.372	59	60.0
7	5-Acetyltryptamine $R_5 = H_3CCO$	8.013	-0.960	-0.0068	-0.0794	11.18	2.689	1.468	63	54.3
8	5- Oxytryptamine (serotonin) R_5 = OH	7.835	-0.677	-0.2229	-0.3197	2.85	2.720	1.515	55	43.3
9	5- Methyltryptamine $R_5 = CH_3$	7.771	-0.599	-0.0018	-0.0245	5.65	2.519	1.297	54	38.1
10	5- Phenyltryptamine $R_5 = C_6H_5$	7.593	-1.098	-0.0113	-0.1166	25.36	2.647	1.264	50	10.0
11	5- Phenyloxytryptamine $R_5 = C_6H_5O$	7.531	-1.123	-0.0105	-0.1149	27.68	2.743	1.415	50	0
12	Tryptamine	7.723	-0.614	-0.0079	-0.0548	1.03	2.583	1.325	50	23.0
13	7- Methyltryptamine $R_7 = CH_3$	7.740	-0.556	-0.0088	-0.0520	1.03	2.519	1.297	54	10.0
14	1- Methyltryptamine R_1=CH_3	7.701	-0.562	-0.0081	-0.0510	1.03	2.519	1.297	10	10.0
15	2- Methyltryptamine $R_2 = CH_3$	7.753	-0.546	-0.0086	-0.0510	1.03	2.519	1.297	82	10.0
16	6- Methyltryptamine $R_6 = CH_3$	7.752	-0.567	-0.0070	-0.0480	1.03	2.519	1.297	59	3.0
17	7- Methyltryptamine $R_7 = OCH_3$	7.629	-0.614	-0.0054	-0.0447	1.03	2.556	1.372	59	4.8
18	4- Methyltryptamine $R_4 = CH_3$	7.663	-0.528	-0.0033	-0.0300	1.03	2.519	1.297	54	35.0
19	7- Chlorotryptamine R_7 = Cl	7.913	-1,053	-0.0005	-0.0245	1.03	2.833	1.325	61	0
20	Tetra-(4,5,6,7)-methyltryptamine $R_4 = R_5 = R_6 = R_7 = CH_3$	7.228	-0.286	0.0041	-0.0173	5.65	2.282	1.199	70	10.0
21	4-Methoxytryptamine $R_4 = OCH_3$	7.400	-0.474	0.0125	0.0529	1.03	2.556	1.372	59	6.0
22	6- Oxytryptamine R_6 =OH	7.765	-0.677	0.0077	0.0592	1.03	2.770	1.515	55	0
23	6- Methoxytryptamine $R_6 = OCH_3$	7.754	-0.637	0.0194	0.0889	1.03	2.556	1.372	59	3.2
24	6- Chlorotryptamine R_6 = Cl	7.980	-1.025	0.0166	0.1323	1.03	2.833	1.325	64	15.0
25	4- Chlorotryptamine R_4 = Cl	7.766	-1.100	0.0159	0.1366	1.03	2.833	1.325	60	0

At the same time, the energy of the highest occupied MO ε_{hoc} turned out to be uninformative for the purpose of interpreting the variability of the radioprotective effect of drugs. This conclusion does not contradict the known results [11] that the value of the energy ε_{hoc} is in no way related to the reactivity of indoles. In addition, it can be noted that the greater the absolute value of the negative charge q_5 on the carbon atom, the higher the ability of the atom to electrophilic attack.

The effect of substituents in the indole ring on the electronic structure of the tryptamine molecule is

reduced to such a distribution of electron density for which the charge on even carbon atoms is positive (or, if negative, it is very insignificant). The charge is usually negative (almost always for effective drugs) on odd atoms (including the fifth carbon atom of the indole ring).

This alternation (oscillation) of the charge is characteristic of the mesomeric effect. The addition of, for example, a chlorine atom to the indole ring sequentially at positions 4, 6, or 7 leads to a change in the value of the complex electronic feature Q_5 and even to a change in its sign on the fifth carbon atom: 0.132, 0.132, -0.025 arb. units, respectively. Such a change in the Q_5 feature is accompanied by a decrease in the antiradiation activity of the drug. At the same time, if the chlorine atom is in the fifth position, then the value of the complex trait increases (in absolute value) to the value Q_5 = -0.457 arb. units, and the antiradiation activity increases markedly. Variations in the distribution of electron density, as well as in the position on the energy scale of the MO energy level ε_{unoc}, have a significant effect on the antiradiation activity of the drug (Table 1). The lower the ε_{unoc} level on the energy scale is, the greater (in absolute value) the negative sign of Q_5 and the higher the antiradiation activity of the substance. A similar situation is also observed, for example, for methoxytryptamine upon substitution at positions 4, 6, and 7 of the indole ring. The following values of the sign were obtained Q_5 = 0.053, 0.089 and -0.045 arb. units, respectively. At the same time, the introduction of the methoxy group in the fifth position increases the negative value of the complex trait: Q_5 = -0.277 arb. units, while the radioprotective activity of the drug sharply increases. That is, the nature of the change in the Q_5 is similar to the situation if the substituent is a chlorine atom (see Table 1). However, in quantitative terms, the Q_5 values differ significantly. As shown by a comparative analysis, for all analyzed substituents, there is the following qualitative relationship - the greater (in absolute value) the negative complex sign Q_5 and the higher the acceptor properties of the substituent R_5 (Table 2), the higher the radioprotective properties of the drug.

Experiment [2] indicates that the simultaneous addition of methyl at positions 4, 5, 6 and 7 of the indole ring reduces the radioprotective activity in comparison with 5-methyltryptamine (Table 1). Indeed, from the data in Table 1 it follows that the introduction of additional substituents reduces the value of the negative sign Q_5 (in absolute value). So, for 5-methyltryptamine, Q_5 = -0.1245 arb. units were determined, and for tetra-(4,5,6,7)-methyltryptamine Q_5 = -0.0173 arb. units, respectively. In addition, the calculation results presented in Table 1 also indicate an unfavorable effect of the substituent on the Q_5 if the substituent is in position 4 (No.17) (compare with No.2 and No.7) and in position 6 (No. 21-24). Moreover, for chemical compounds No. 21-24, the Q_5 changes sign (becomes positive), which does not enhance the radioprotective effect of drugs. Apparently, the following circumstance is of great importance, namely, the effect of the substituent in the indole ring on the value of the quantum feature Q_5. For example, 7-chlorotryptamine has no radioprotective effect, but at the same time gives a negative value Q_5 = -0.025 arb. units. However, additional the accession of the methoxy group (No. 2) at position R_5 leads to a significant change in the value of the feature Q_5 = -0.4866 arb. units. In absolute value, this value significantly exceeds Q_5 = -0.2767 arb. units (Table 1) for 5-methoxytryptamine (No. 6).

Quantum-mechanical calculations of the electronic structure of molecules are not always available to researchers and require a qualified analysis of the results of calculations. Therefore, it is of certain interest to find a computationally simple method that allows one to independently obtain an estimate of the value of the feature Q_5. For some substituents, their electronic affinity values (A, in units of eV) are known [12]. The observed values of the electron affinity of atomic groups are shown in Table 2. We can note the parallelism between the value of the electronic attribute Q_5 calculated by the quantum-chemical method and the value of the electron affinity of the substituent R_5 (Table 2). Affinity energy is often used to quantify the acceptor properties of electronic systems.

The factor signs Q_{5i} and A_i are assumed to be independent realizations of a pair of random variables (Q_5, A > 0). The linear correlation equation (Fig.2A) between the value of the molecular sign Q_5 and the value of the electron affinity A of the substituents (Table 2) has the following statistics:

$Q_5(A) = a_0 + a_1 \cdot A$, $n = 10$, $R = -0.956 \pm 0.03$, $R^* = 0.97 > R_{0.05}^{cr}(n - 2) = 0.632$; sample size sufficient for the validity of the correlation coefficient: $n_{0.05}^{min} < 5$ [15]; $a_0 = 0.015 \pm 0.036$, $a_1 = -0.142 \pm 0.015$, $|t(a_1)| = 9.3 > t_{0.05}^{cr}(n - 2) = 2.306$; $RMSE = 0.051$; $F = 87.04 > F_{0.05}^{cr}(f_1 = 1; f_2 = 8) = 5.32$; sum of squares residuals: $\Sigma = 0.0204$; straightforwardness: $K = (N \cdot (1 - R^2))^{0.5} = 0.92 < K^{thr} = 3.00$ [15].

(1)

Here *RMSE* (Root Mean Square Error) is the standard error of the estimate (i.e., the measure of variation of an effective trait with respect to the regression line). The *RMSE* value indicates how much the regression (1) allows the model to approximate the original data. For a sample size $N \leq 50$, it is preferable to use the corrected linear correlation coefficient: $R^* = R\cdot[1 + 0.5\cdot(1 - R^2)/(N - 3)]$.

Typically, a model is considered acceptable if the coefficient of determination is greater than 0.5. A more detailed assessment of model quality is carried out using critical conditions (1). Comparing the empirical values of the statistics with the tabular values, we come to the conclusion about the significance of the estimates at the level of $\alpha = 0.05$. The ratio $F = R^2\cdot (n - m - 1) / (1 - R^2)$ [8,16] significantly exceeds the tabular value of F^{cr}, therefore, there are reasons for recognizing the importance of the relationship between signs Q_5 and A; m is the number of explanatory variables. For a sample size $n \geq 10$, the significance of the correlation coefficient is checked using the following inequality:

$$t = |R|\cdot(n - 2)^{0.5}/(1 - R^2)^{0.5} = 9.7 > t_{0.05}^{cr}(n - 2) = 1.86. \quad (2)$$

Table 2

The values of the electron affinity of the substituents, factor Z (arb. units), the value of the Q_5 trait (arb. units), the observed survival Act_{exp} (%) and the dose of the drug (mg/kg)

R_5	Cl	F	Br	J	OH	OCH_3	OC_2H_5	CH_3	OC_6H_5	H
Q_5 *)	-0.457	-0.504	-0.427	-0.444	-0.320	-0.277	-0.181	-0.125	-0.115	-0.055
A, eV[12]	3.61	3.45	3.37**)	3.08*)	1.83	1.50	1.60**)	1.10	1.20**)	0.75
Z_{sub}	7	7	7	7	3.50	2.60	2.38	1.75	2.92	1.00
Z***)	2.833	2.833	2.833	2.833	2.720	2.556	2.581	2.519	2.743	2,583
Act_{exp}	68	67.5	58	59.9	43.3	60	16	38.1	0	23
Dose	60	56	74	89	55	59	75	54	50	50

*) The Q_5 values are calculated by the quantum mechanical method. **) The values of the electron affinity were obtained by comparing various data [13,14]. ***) For halogen atoms, it is assumed that the outer electron shell has the configuration s^2p^5.

For sample sizes $N < 50$, the following estimate of the significance of the linear correlation coefficient is used, based on the normalizing Fisher z-transform, taking into account Hotelling's corrections for the sample size [16,18]:

$$u_H = 1.91 > u_{0.05}(N) = z_{0.975}\cdot(N - 1)^{-0.5} = 0.653. \quad (3a)$$

For the correlation coefficient, the boundaries of the confidence interval at a significance level of $\alpha = 0.05$ are determined by the inequality:

$$R_1 = 0.850 < R^* < R_2 = 0.988,$$

$$R_{1,2} = (\exp(2\cdot u_{1,2}) - 1)/(\exp(2\cdot u_{1,2}) + 1),$$

$$u_1 = u - u_{0.05}(N), \qquad u_1 = u + u_{0.05}(N). \quad (3b)$$

Here, for the quantile of the normal distribution, it is assumed that $z_\alpha = 1.960$ at $\alpha = 0.975$. The linear correlation coefficient can be considered statistically significant since the empirical value $u_H > u_\alpha(N)$. If the empirical value u satisfies the inequality $|u| < u_\alpha(N)$, then the null hypothesis is accepted. That is, there is no relationship between the quantities under consideration. Therefore, both criteria (2) and (3*a*) do not contradict each other. Thus, at a significance level of 5%, one can assume the presence of a "very close" (according to the Chaddock scale [19] linear correlation between the features Q_5 and A.

Regression (1) explains 94% of the variability of the effective sign and only 6% (the uncertainty coefficient is 0.06) is due to some unaccounted for factors or random scatter of data. The high significance and predictive [20] ability of regression is indicated by the value of the ratio F, which is significantly higher than the table value.

Since the regression coefficient $a_1 < 0$ (2) is significant, there is a statistically significant (at the 95% confidence level) negative linear relationship between the effective indicator $Q_5 < 0$ and the explanatory variable A > 0. It is also possible to check how sufficient it is to take into account only the linear contribution A in the regression equation (1). Let us check whether it is necessary to additionally take into account, for example, the quadratic contribution A^2. To do this, check the relationship between the regression residuals (1) and the value of A^2. The correlation coefficient was

found to be insignificant (R = 0.05). Therefore, the additional explanatory variable is not significant. It can also be noted that, in accordance with regression (1), the complex molecular parameter should vanish at the electron affinity A = 0.

A statistically significant relationship was also found between the sign Z_{sub} which characterizes the pseudopotential of the atomic group (Table 2), and the value of the explanatory variable Q_5:

$Q_5(Z_{sub}) = a_0 + a_1 \cdot Z_{sub}$, n = 10, R = -0.95 ± **0.03**, $|R^*|$ = 0.96 > $R_{0.05}^{cr}(n - 2)$ = 0.632; sample size sufficient for the validity of the correlation coefficient: $n_{0.05}^{min}$ < 5; a_0 = - 0.025 ± 0.037, a_1 = - 0.064 ± 0.008, $|t(a_1)|$ = 8.42 > $t_{0.05}^{cr}(n - 2)$ = 2.306; criterion for the significance of the correlation coefficient based on the Fisher normalizing z-transform (taking into account Hotelling's corrections [16]): u_H = 1.776 > $u_{0.05}(n) = z_{0.975} \cdot (n - 1)^{-0.5}$ = 0.653; $RMSE$ = 0.056; F = 70.85 > $F_{0.05}^{cr}(f_1 = 1; f_2 = 8)$ = 5.32; sum of squares residuals: Σ = 0.0255; statistical features of straightness: $K = (N \cdot (1 - R^2))^{0.5}$ = 0.99 < K^{thr} = 3.00.

(4)

Sample statistics for sets Q_5, A, and Z_{sub}:

Q_5: n = 10; Q_5^{av} = - 0.29 ± 0.05; 95% confidence interval: (-0.41, -0.17); Q_5^{min} = -0.504, Q_5^{max} = - 0.055, S_Q = 0.164; τ^{min} = 1.30 < τ^{max} = 1.44 < $\tau_{0.05}^{cr,2}(n)$ = 2.294< $\tau_{0.05}^{cr,1}(n)$= 2.441; Wilk-Shapiro normality test: W = 0.914 > $W_{0.05}^{cr}(n)$= 0.842, David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(n)$= 2.670 < $U = [(Q_5^{min} - Q_5^{max})/S_Q]$ = 2.74 < $U2_{0.05}^{cr}(n)$ = 3.685;

A: n = 10; A^{av} = 2.15 ± 0.35; 95% confidence interval: (1.36 - 2.94); A^{min} = 0.75, A^{max}= 3.61; S_A = 1.164; τ^{max} = 1.26 < τ^{min} = 1.20 < $\tau_{0.05}^{cr,2}(n)$ = 2.294 < $\tau_{0.05}^{cr,1}(n)$= 2.44; Wilk-Shapiro normality test: W = 0.866 > $W_{0.05}^{cr}(n)$= 0.842, David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(n)$= 2.67 ≈ $U = [(A^{max} - A^{min})/S_A]$ = 2.46 < $U2_{0.05}^{cr}(n)$ = 3.685;

Z_{sub}: n = 10; Z^{av} = 4.22 ± 0.79; 95% confidence interval: (2.44 - 5.99); Z^{min} = 1.00, Z^{max}= 7.00, S_Z = 2.486, τ^{max} = 1.12 < τ^{min} = 1.29 < $\tau_{0.05}^{cr,2}(n)$ = 2.294< $\tau_{0.05}^{cr,1}(n)$ = 2.44; Wilk-Shapiro normality test: W = 0.813 ≈ $W_{0.05}^{cr}(n)$= 0.842, David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(n)$= 2.670 ≈ $U = [(Z^{max} - Z^{min})/S_Z]$ = 2.414 < $U2_{0.05}^{cr}(n)$ = 3.685.

(5)

The following remark should be made here. In table 2 for the same values Z_{sub} = 7.0 arb. units (Nos. 1 - 4) different values of Q_5 are given. In this case, one average Q_5^{av} can be used instead of separate Q_5 values for each Z_{sub}. Using averages allows you to align different sample values by showing the average level of the characteristic values. In this case, the use of averages does not lead to significant changes in further results obtained from grouped data.

The sets Q_5, A and Z_{sub} are homogeneous and have a distribution close to the normal distribution. The homogeneity of the samples is checked using the Grubbs-Romanovsky τ-test at the significance level α = 0.05 and the sample size n. The test assumes that for a homogeneous sample, the extreme values of the population are less than the table values. The reliability of the difference of the regression coefficients from zero is estimated by comparing their t-value with the table value: $t_{0.05}^{cr}(f = n - m - 1)$; here f is the number of degrees of freedom and the confidence level is 95%; m is the number of links. If the empirical value of t is greater than the tabular value at the significance level α, then it can be assumed that the regression coefficient a_1 is reliable with a probability of $1 - \alpha$.

The analysis of the regression residuals (1) showed that, apparently, the initial data are not stable. To check the stability of the series, one can use the Chow test [21]. For this, the series under study (Table 2) is proposed to be divided into two subsamples. The first subsample is bioactive chemical compounds (Nos. 1 - 6). The second subsample contains low-active drugs (nos. 7 - 10). The possibility of combining two samples into one sample is determined using the Chou test [21]. A null hypothesis is put forward about the structural stability of the trend of the analyzed series (1).

For each of the subsamples, its own linear regression equation is constructed (Fig. 2A) and the sum of the squares of the regression residuals is determined: Σ_1 = 0.0032 (first subsample; n_1 = 6; No. 1-6) and Σ_2 = 0.00034 (second subsample; n_2 = 4; Nos. 7-10). The sum of the squares of the residuals for the general regression is Σ = 0.0204 ($n = n_1 + n_2$ = 10). The regression residuals are normally distributed. The difference $\Sigma - \Sigma_1 - \Sigma_2$ characterizes the improvement in the quality of the model after dividing the combined sample (1) into two subsamples. The distribution of the residuals for each of the regressions satisfies the normal distribution. Then the relation that has the F-distribution is calculated:

$$F = (\Sigma - \Sigma_1 - \Sigma_2)(n - 2m - 2)/(\Sigma_1 + \Sigma_2)/(m + 1)$$

$$= 14.16 >$$

$$F_{0.05}^{cr}(f_1 = m + 1; f_2 = N - 2m - 2) = 5.14. \quad (6)$$

Here $m = 1$ is the number of explanatory variables. Since the ratio F is significantly greater than F^{cr}, it can be recognized that the variation series is unstable. Consequently, the regression analysis of the individual independent parts of the sample gives a better result (the significance of structural changes) than the regression analysis of the entire sample as a whole. It follows from inequality (6) that the null hypothesis should be rejected and therefore two regressions should not be combined into a single linear regression (1). That is, it is fair to split the set of observations ($n = 10$) into two parts, and each of the parts must be approximated by its own linear dependence. Such a discontinuity (approximately in the range of $A^{thr} \approx 1.5$ - 1.6 eV) in the continuity of the interrelation of molecular properties can be associated with a structural shift in the interrelationships of the explanatory variables. A structural shift separates bioactive drugs from weak and inactive chemical compounds.

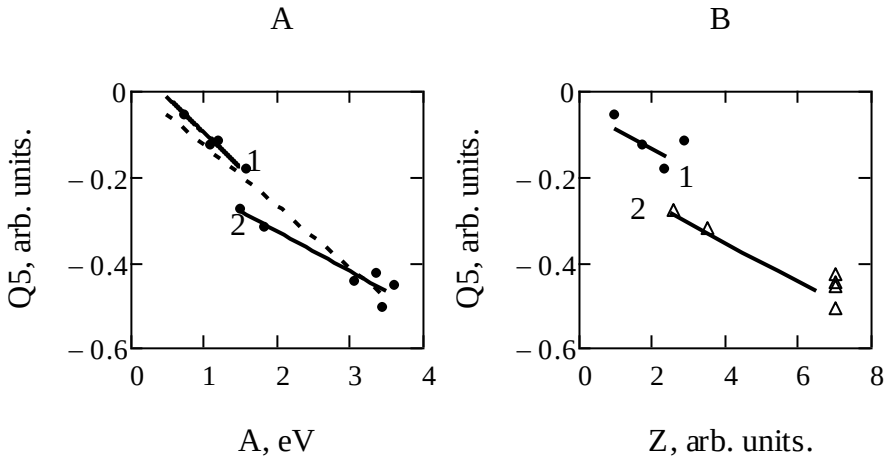

Fig. 2. Scatter plots. A. The relationship between the Q_5 feature and the energy of affinity for the electron A of the R_5 substituent. The dashed regression line is given by equation (1). Weakly active or inactive drugs (Nos. 7-10; Table 2) - line 1: the regression $Q_{51}(A) = 0.049 – 0.145 \cdot A$, $n_1 = 4$, $\Sigma_1 = 0.000336$. Bioactive drugs (Nos. 1-6; Table 5.2) - line 2: the regression is: $Q_{52}(A) = - 0.146 – 0.092 \cdot A$, $n_2 = 6$, $\Sigma_2 = 0.00323$.
B. The relationship between the feature Q_5 and the electronic factor of the substituent $Z_{sub}(R_5)$. Weakly active or inactive drugs (Nos. 7-10; Table 2) - line 1: the regression is $Q_{51}(Z) = - 0.038 – 0.040 \cdot Z$, $n_1 = 4$, $\Sigma_1 = 0.0047$. Bioactive drugs (Nos. 1-6; Table 2) - line 2: the regression is $Q_{52}(Z) = - 0.174 – 0.041 \cdot Z$, $n_2 = 6$, $\Sigma_2 = 0.0033$.

A similar analysis was performed for the relationship between the signs Q_5 and Z_{sub} (Table 2). Again, there is a structural shift (Fig. 2B) between the explanatory variables. The following inequality was obtained for the Chow test:

$$F = 7.09 > F_{0.05}^{cr}(f_1 = m + 1; f_2 = N - 2m - 2) = 5.14. \quad (7)$$

The boundary value of the factor Z^{thr} is in the range of 2.5 - 2.6 arb. units. Thus, according to the Chow test, the samples for active and weakly active compounds differ. Therefore, it is not recommended to combine them into a single lenear regression. For the substituents shown in Table 2, there is a significant relationship between the value of the electron affinity A of the substituent and the value of the molecular characteristic Z_{sub}:

$A(Z_{sub}) = a_0 + a_1 \cdot Z_{sub}$, $n = 10$, $R = 0.98 \pm 0.01$, $|R^*| = 0.99 > R_{0.05}^{cr}(n - 2) = 0.632$; sample size sufficient to validate the correlation coefficient: $n_{0.05}^{min} < 5$; $a_0 = 0.31 \pm 0.14$, $a_1 = 0.44 \pm 0.03$, $t(a_1) = 15.3 > t_{0.05}^{cr}(n - 2) = 2.306$; criterion for the significance of the correlation coefficient based on the Fisher normalizing z-transform (taking into account Hotelling's corrections): $u_H = 2.424 > u_{0.05}(n) = z_{0.975} \cdot (n - 1)^{-0.5} = 0.653$; $RMSE = 0.213$; $F = 233.6 > F_{0.05}^{cr}(f_1 = 1; f_2 = 8) = 5.32$; statistical sign of straightness: $K = 0.45 < K^{thr} = 3.00$. (8)

It follows from regression (8) that an increase in the molecular characteristic Z_{sub} is associated with an increase in the value of A (there is a positive linear

regression). That is, when an electron is attached, the released energy (A > 0) increases.

Comparative data presented in Table 2 allow, firstly, to make an approximate quantitative assessment of the total inductive and mesomeric effects of substituents in aromatic compounds. Secondly, it provides an approximate rapid assessment of the radioprotective bioactivity of 5-substituted tryptamine derivatives without performing cumbersome and laborious quantum mechanical calculations. In addition, regression (1) implies a clear physical meaning of the complex molecular characteristic Q_5. Molecular trait Q_5 characterizes two possible processes in which a drug molecule can participate. First, in directed electrostatic binding with the target object of the biosystem due to the electrostatic interaction of a negative charge in the region of the carbon atom C_5. Secondly, it changes the acceptor properties of the molecule. The higher the MO energy level is on the energy scale $\varepsilon_{unoc} > 0$, the weaker the acceptor properties of the molecule as a whole. At the same time, in accordance with the data in Table 2 and regression (1), the radioprotective properties of drugs increase with an increase in the acceptor properties of the R_5 substituent. However, tryptamine derivatives can also have donor properties. For example, according to A.Szent-Györgyi [11], serotonin should be a good univalent electron donor. In this case, presumably (since there is no information about the MO electron energies of the acceptor), a stable molecular complex (dimer) can form, in which the interaction of molecules is due to electron transfer. Indeed, the explanatory factor *Z*, associated with the number of valence electrons of the R_5 substituent, is associated with the bioactivity of the drugs (Table 2). It should also be noted that indole derivatives have a very high reactivity, not necessarily associated only with the donor-acceptor properties of molecules.

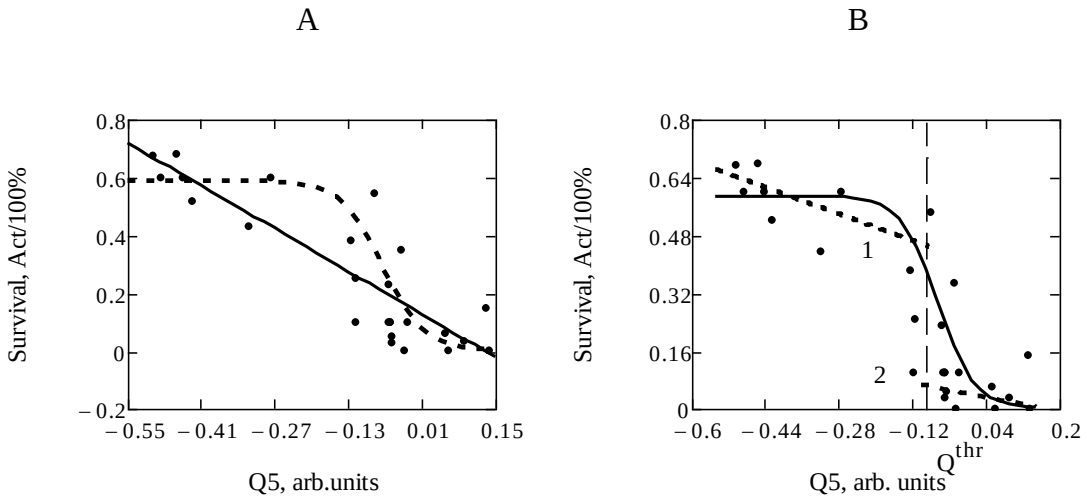

Fig. 3. Changes in the radioprotective effect of 5-substituted tryptamines depending on the magnitude of the quantum feature Q_5. A. The solid line is defined by equation (9). The dotted line indicates the putative threshold of biological action. B. Dotted line (1): $Act1/100 = 0.38 – 0.64 \cdot Q_5$, $N_1 = 10$, $R_1 = - 0.81$, $RMSE = 0.07$; dotted line (2): $Act2/100 = 0.05 – 0.24 \cdot Q_5$, $N_2 = 13$, $R_2 = - 0.42$, $RMSE = 0.04$. The solid nonlinear curve demonstrates the threshold of radioprotective action and is approximated by a nonlinear form: $Act_{nonl}/100 = [1.70 + 1.27 \cdot \exp(1.90 + 25.0 \cdot Q_5)]^{-1}$, $RMSE = 0.165$, $N = 25$. $Q_5^{thr} \approx - 0.09$ усл. ед. The vertical dashed line indicates the boundary value of the factor $Q_5^{thr} \approx -0.09$ arb. units.

Figures 3A and 3B show scatter diagrams of the observed radioprotective effect of drugs for a group of compounds (Nos. 1–25; Table 1). A scatter plot allows visual analysis of empirical data. The scatterplot suggests which single regression function can be used to approximate the relationship between the explanatory variable and the effective indicator.

For example, the following equation of linear correlation was obtained (Fig. 3A), which

determines the dependence of bioactivity on the value of the explanatory variable Q_5:

$Act3/100 = a_0 + a_1 \cdot Q_5$, $N = 25$, $R = -0.86 \pm 0.05$, $|R^*| = 0.87 > R_{0.05}^{cr}(N - 2) = 0.396$; $RMSE = 0.132$; sample size sufficient for the validity of the correlation coefficient: $N_{0.05}^{min} = 5$; $a_0 = 0.12 \pm 0.03$, $a_1 = -1.11 \pm 0.14$, $|t(a_1)| = 8.2 > t(a_0) = 3.7 > t_{0.05}^{cr}(N - 2) = 2.06$; unexplained regression residuals (disturbing variable) are normally distributed: Wilk-Shapiro test: $W = 0.955 > W_{0.05}^{cr}(N) = 0.918$; $F = 67.0 > F_{0.05}^{cr}(f_1 = 1; f_2 = 23) = 4.28$; $\Sigma = 0.398$. (9)

Since the empirical value is $F >> F^{cr}$ the Q_5 variable at the significance level $\alpha = 0.05$ reliably explains the variability of bioactivity. That is, 74% of the total variance in bioactivity is due to a change in the explanatory variable Q_5.

Statistics of sets Q_5 and $Act3/100$:

$N = 25$, $Q_5^{av} = -0.13 \pm 0.04$; 95% confidence interval: (-0.21, -0.05); $Q_5^{min} = -0.50$, $Q_5^{max} = 0.137$, $S_Q = 0.20$, $\tau^{max} = 1.34 < \tau^{min} = 1.85 < \tau_{0.05}^{cr,2}(N) = 2.717 < \tau_{0.05}^{cr,1}(N) = 2.815$; Wilk-Shapiro normality test: $W = 0.872 < W_{0.05}^{cr}(N) = 0.918$, David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(N) = 3.34 \approx U = [(Q_5^{max} - Q_5^{min})/S_Q] = 3.19 < U2_{0.05}^{cr}(N) = 4.71$;

$N = 25$, $Act_{exp}^{av}/100 = 0.27 \pm 0.05$; 95% confidence interval: 0.16-0.37; $Act_{exp}^{min}/100 = 0$, $Act_{exp}^{max}/100 = 0.693$, $S_{Act3} = 0.261$, $\tau^{min} = 1.03 < \tau^{max} = 1.62 < \tau_{0.05}^{cr,2}(N) = 2.717 < \tau_{0.05}^{cr,1}(N) = 2.815$; Wilk-Shapiro normality test: $W = 0.828 < W_{0.05}^{cr}(N) = 0.918$, David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(N) = 3.34 > U = [(Act_{exp}^{max}/100 - Act_{exp}^{min}/100)/S_{Act3}] = 2.65 < U2_{0.05}^{cr}(N) = 4.71$.

(10)

Sets of elements are homogeneous and have a distribution close to normal distribution.

The regression equation (9) only indicates the following relationship: the more negative the value of the Q_5 trait, the more likely the higher the radioprotective effect of the drug. This is also indicated by the Abbe-Linnik statistical test [18,23]:

$$q = 0.5 \cdot \sum_{i=1}^{N-1} \left(Q_{5,i+1} - Q_{5,i}\right)^2 / \sum_{i=1}^{N} \left(Q_{5,i} - Q_5^{av}\right)^2 = 0.09 < q_{0.05}^{cr}(N) = 0.6839.$$

The choice of the regression equation in linear form was arbitrary. Therefore, an arbitrary choice of the analytical form of the regression equation can lead to an incorrect interpretation of the analyzed relationship. It is necessary to perform an additional statistical check, namely, to find out whether the initial data of the bioresponse as a whole ($N = 25$) can be interpreted as a linear dependence on the Q_5 trait. Thus, is it possible to expect a significant increase in radioprotective activity with an increase (in absolute value) of the negative complex feature Q_5.

Let us analyze the residuals δA of regression (9) (residuals are normally distributed: Wilk-Shapiro test $W = 0.934 > W_{0.05}^{cr}(N = 25) = 0.918$). Further, using a polynomial of the fourth degree as an approximate trial approximating function, it was found that the function characterizing the relationship between the residuals and the value of the feature Q_5 has an inflection in the region of the value of the feature $Q_5 \approx -0.12$ arb. units. As you know, at the inflection point, the first derivative of a function changes its feature. This Q_5 value can be roughly taken as a threshold value. That is, with this value of the explanatory variable Q_5, in the area of which there is a significant change in the bioactivity of drugs (the area of transition from weak bioactivity to relatively high). The branches of the approximating curve to the left and to the right of the inflection point are in areas of relatively high and low bioactivity, respectively (Fig. 4).

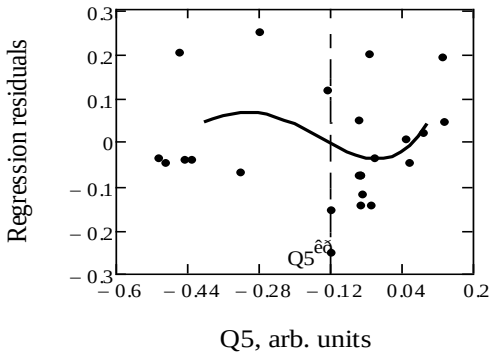

Fig. 4. Scatter plot of the regression residuals (9). The dashed vertical line identifies the inflection point of the regression curve approximated by the following polynomial: $\delta A(Q_5) = -0.04 + 0.11 \cdot Q_5 + 5.2 \cdot Q_5^2 + 15.8 \cdot Q_5^3 + 12.0 \cdot Q_5^4$.

The series of residues $\delta A_i = A_{exp} - A_{mod}$ regression (9) (here $A_{exp} = Act_{exp}/100$ and $A_{mod} \equiv Act3/100$) has the following sequence (from higher bioactivity to lower):

-0.0381, -0.0490, 0.2017, -0.0438, -0.0424, 0.2479, 0.3310, -0.0728, 0.1158, -0.1559, -0.2539, 0.0471, -0.0796, -0.0784, -0.0784, -0.1449, -0.1229, 0.1964, -0.1471, -0.0386, 0.0043, -0.0484, 0.0189, 0.1881, 0.0432 (11)

Let's check the series (11) for randomness and, thus, for the absence of a trend. For this, we use the criterion for the number of series of signs of the first differences [18]. Let us successively calculate the differences $\Delta = \delta A_{i+1} - \delta A_i$. If $\Delta > 0$, then this difference is assigned the index "+", if $\Delta < 0$, then we assign the index "-", if the differences are equal, then we put the index "0". Using series (11), the following sequence of series of indices was obtained:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
- + - +++ - + -- + - + 0 - ++ - ++ - ++ - (12)

This sequence of characters includes **R** = 18 series, with an initial sample size of N = 25. Series numbers are indicated by ciphers (12). Then the quantity of series is compared with the critical values $\mathbf{R}_1$ (lower limit) and $\mathbf{R}_2$ (upper limit). At the 95% confidence level, we obtain the following inequalities: $\mathbf{R}_1(\alpha = 0.05) = 11 < \mathbf{R} = 18 < \mathbf{R}_2(\alpha = 0.05) = 21$, which indicate the randomness of sequences (11) and (12). Additional information about the randomness of the residuals can be obtained by evaluating the mean value $M(\mathbf{R}) = (2 \cdot N - 1)/3 = 16.33$ and variance $D(\mathbf{R}) = (16 \cdot N - 29)/90 = 4.12$. The randomness hypothesis is tested by comparing the value $|\mathbf{R}^*| = |\mathbf{R} - M(\mathbf{R})|/D(\mathbf{R}) = 0.823 < u_{0.975} = 1.96$ with the normal distribution quantile $u_{1-\alpha/2}$ at $\alpha = 0.05$. Thus, this criterion also indicates the randomness of a number of residuals and the absence of a trend. Consequently, linear regression (9) generally reveals a trend linking the value of the electronic factor Q_5 and the survival rate of experimental animals.

The Chow test (5) can be used again to determine the stability of the whole analysed series. For this, the series under study (Table 1) is divided into two subsamples. The first subsample (Nos. 1-7) contains drugs that have a radioprotective activity $Act > 50\%$, and the second (Nos. 8-25) is characterized by a relatively low bioactivity $Act < 45\%$. A null hypothesis is put forward about the structural stability of the trend of the series (9). For each of the subsamples, a linear regression equation is written (Fig. 3B) and the sum of the squares of the regression residuals is determined: $\Sigma_1 = 0.0158$ (first subsample; $N_1 = 7$) and $\Sigma_2 = 0.2092$ (second subsample; $N_2 = 18$). The sum of the squares of the residuals for the general regression is $\Sigma = 0.4868$ ($N = N_1 + N_2 = 25$). Next, the ratio is calculated, which has the F-distribution:

$$F = (\Sigma - \Sigma_1 - \Sigma_2)(n - 2m - 2)/(\Sigma_1 + \Sigma_2)/(m + 1)$$

$$= 12.22 > F_{0.05}^{cr}(f_1 = m + 1; f_2 = N - 2m - 2) = 3.47. \quad (13)$$

Here $m = 1$ is the number of explanatory variables. Since $F > F^{cr}$ therefore the variation series is unstable. It follows from inequality (13) that the null hypothesis should be rejected and therefore it is not recommended to combine two regressions into one linear regression. Moreover, this result also suggests that, for the combined sample (N = 25), the relationship between the bioactivity and the explanatory variable should be defined by a broken line. This is demonstrated in Figs. 3B and 4. The existence of two different linear dependencies (Fig. 3B) for bioactive drugs (line 1) and for low-activity (line (2)) drugs indicates the presence of a structural shift in the relationship of molecular traits.

Let's also perform an additional check for nonlinearity of the relationship $Act(Q_5)/100$ for the combined sample (N = 25). We will use the technique described in [24,25]. We preliminary rank the chemical compounds from Table 1 by the Q_5 attribute. Next, we divide the entire set into approximately homogeneous groups. The number of groups can be determined using the H.A.Sturges ratio [24]: $n \approx 1+3.32 \cdot \lg(N) = 6$. The width of the intervals of the groups can be approximately assumed to be the same. Using the ratio: $\Delta Q_5 \approx (Q_5^{max} - Q_5^{min})/n$, we determine the width of the interval (0.11 - 0.13) arb. units. Here Q_5^{max} and Q_5^{min} are the maximum and minimum values of the complex attribute Q_5. Thus, six groups were formed, which included 4, 2, 2, 2, 13 and 2 drugs. For each group, the mean values were calculated $Q_{5,i}^{av} \equiv Q_I^{av}$, $Q_{II}^{av}, \ldots, Q_{VI}^{av}$, and $Act_i^{av}/100 \equiv Act_I^{av}, \ldots, Act_{VI}^{av}$. Then the b_{i-j} ratios of the compared groups were determined:

$$b_{I\text{-}II} = (Act_I^{av} - Act_{II}^{av})/(Q_I^{av} - Q_{II}^{av}) = -1.63,$$

$$b_{I\text{-}III} = (Act_I^{av} - Act_{III}^{av})/(Q_I^{av} - Q_{III}^{av}) = -0.54,$$

$$b_{III\text{-}IV} = (Act_{III}^{av} - Act_{IV}^{av})/(Q_{III}^{av} - Q_{IV}^{av}) = -5.21,$$

$$b_{IV\text{-}V} = (Act_{IV}^{av} - Act_V^{av})/(Q_{IV}^{av} - Q_V^{av}) = 0.80. \quad (14)$$

Here i, j = I, II, III, IV, V and VI are group numbers. For a linear relationship, the ratios (14) must be close in magnitude. However, the obtained differences between the b_{i-j} values are significant. Even the sign of the attitude changes. Therefore, a series of ratios (14) for the entire sample (N = 25)

indicates that the relationship between the explanatory variable Q_5 and bioactivity must be either non-linear or there is a structural break (Fig. 3B). Thus, the linear relationship between the bioresponse and the explanatory variable for the entire sample (Fig.3A), apparently, should be abandoned. The angle of inclination of the line (9) relative to the abscissa axis is significantly greater than the angle of inclination for line (1) of the figure (Fig. 3B). Considering the nonlinearity of the relationship between the change in molecular structure and the variability of the bioactivity, a significant linear increase in the bioactivity of drugs with an increase in the feature Q_5, apparently, cannot be expected.

Here i, j = I, II, III, IV, V and VI are group numbers. For a linear relationship, the ratios (14) must be close in magnitude. However, the obtained differences between the b_{i-j} values are significant. Even the sign of the attitude changes. Therefore, a series of ratios (14) for the entire sample (N = 25) indicates that the relationship between the explanatory variable Q_5 and bioactivity must be either non-linear or there is a structural break (Fig. 3B). Thus, the linear relationship between the bioresponse and the explanatory variable for the entire sample (Fig.3A), apparently, should be abandoned. The angle of inclination of the line (9) relative to the abscissa axis is significantly greater than the angle of inclination for line (1) of the figure (Fig. 3B). Considering the nonlinearity of the relationship between the change in molecular structure and the variability of the bioactivity, a significant linear increase in the bioactivity of drugs with an increase in the feature Q_5, apparently, cannot be expected.

Let us also check for the presence of a fast (jump) change in the effective indicator with a relatively small change in the explanatory variable Q_5. The following two samples are analyzed, ranked according to the value of bioactivity (dimensionless value):

A/100%: 0.68 0.675 0.644 0.60 0.599 0.58 0.433 0.381 0.10 0

Q_5: -0.4572 -0.5036 -0.4866 -0.2767 -0.4441 -0.4269 -0.3197 - 0.1245 -0.1166 -0.1149. (15)

Sequence (15) includes chemical compounds (N = 10), for which $Q_5 < Q_5^{thr}$ = -0.09 arb. units. Next, let's use the Cochran test [16,18]. If in a series of observations the mean value changes abruptly, then the difference between the mean values $A1i^{av} - A2i^{av}$ corresponding to the first observations N_i and subsequent $N - N_i$ observations is checked using χ^2 - statistics (two-sided test with one degree of freedom):

$$\chi^2 = N_i(N - N_i)(A1_i^{cp} - A2_i^{cp})^2/N/A^{cp},$$
$$A^{cp} = N^{-1}\sum_{i=1}^{N} A_i,\ A1_i^{cp} = N_i^{-1}\sum_{i=1}^{N_i} A_i,$$
$$A2_i^{cp} = (N - N_i)^{-1}\sum_{i=N_{i+1}}^{N} A_i\ , \quad N_i + N_{i+1} = N,\ i = 2,3,\ldots,N-1. \tag{16}$$

Here $A1_i^{av}$ is the overall average; $A2_i^{av}$ is the average value corresponding to the first N_i observations; $A1_i^{av}$ is the mean value of the set of subsequent $N - N_i$ observations. As a result, we obtain the following series of quantities for the average values of bioactivity.

Average of the first two values:

$A1_2^{av}$ = (0.68 + 0.675)/2,

average of subsequent eight values in the original series:

$A2_2^{av}$ = (0.644+0.60+0.599+0.58+0.433+

0.381+0.10+0)/8 = 0.4171.

In the same way, we calculate the average values for the subsequent steps:

$A1_3^{av}$ = 0.6663, $A2_3^{av}$ = 0.3846, $A1_4^{av}$ = 0.6498, $A2_4^{av}$ = 0.3487,

$A1_5^{av}$ = 0.6396, $A2_5^{av}$ = 0.2986, $A1_6^{av}$ = 0.6297, $A2_6^{av}$ = 0.2283,

$A1_7^{av}$ = 0.6016, $A2_7^{av}$ = 0.160, $A1_8^{av}$ = 0.5739, $A2_8^{av}$ = 0.50. (17)

Using the mean values (17), we can calculate the values of χ^2-statistics for each subsequent step:

$$\chi_2^2 = 2.313 < \chi_3^2 = 2.78 < \chi_{0.05}^{2,cr}(f = 1) = 3.841 \approx \chi_4^2 = 4.63 < \chi_5^2 = 6.94 < \chi_6^2 = 8.73. \tag{18}$$

Thus, after the fourth step (18), there is a significant increase in the value of χ^2. Such an increase can be compared with a jump in the average statistical series. That is, the dynamics of the relationship of bioactivity depending on the change in the explanatory variable is likely to have a threshold character.

Figure 3B shows a possible non-linear relationship between the explanatory variable Q_5 and the radioprotective effect of drugs. From the form of this dependence it follows: firstly, the feature Q_5 has a threshold value of Q^{thr}, after which there is a sharp change in bioactivity. Secondly, there are two

different areas for drugs: before Q^{thr} and after the threshold value of Q^{thr}. That is, there are two structurally different areas (lines 1 and 2; Fig. 3B), for which significant changes in the value of the molecular trait Q_5 do not lead to very significant changes in the bioactivity of drugs. This conclusion does not contradict Chow's test (13).

If we take approximately for the critical (threshold) value of the attribute Q_5 the value $Q^{thr} \approx$ - 0.09 arb. units (which is very close to the curve bend point (Fig. 4)), then a group of effective radioprotectors covers the area $Q_5 \leq Q^{thr}$ (Fig. 3B; line 1). The greater the value (in absolute value) of the negative parameter Q_5, the higher the biological activity of chemical compounds. Therefore, it can be assumed that the receptor region to which the substituted tryptamine molecule interacts should be positively charged. In addition, it should be noted that the Q^{thr} parameter indicates the boundary after (in Fig. 3, the direction to the right of Q^{thr}) which the bioactivity of the molecules sharply decreases (Fig. 3B; line 2). That is, the radioprotective effect of tryptamines with substituents in the indole ring seems to have a threshold in terms of the magnitude of the complex quantum trait. At the same time, for the range of values $|Q_5| > |Q^{thr}|$ (in the figure, the direction to the left of Q^{thr}), the radioprotective activity of tryptamine derivatives increases rapidly. This situation is possible if the biological effect of the molecule is determined by the strength of its bond with the receptor. Until the bond strength reaches a certain value, the association of the molecule plus receptor complex is unstable. Instability can arise, for example, due to the thermal motion of molecules. In addition, it should be taken into account that the complex electronic feature Q_5 also depends on the position of the energy level ε_{unoc}. The lower (i.e., more in absolute value) this level is on the energy scale, the greater in absolute value the negative value of the Q_5 factor and, therefore, the higher the radioprotective activity of the drug. It is well known, that a lower position on the energy scale of the one-electron level ε_{unoc} creates preferable conditions for the participation of a molecule in electron transfer processes. This does not contradict the data given in Table 2. Taking into account the statistical significance of regression (1), it can also be argued that the stronger the acceptor properties of the substituent in the R_5 position, the higher the protective properties of the drugs. However, in contrast to Fig. 3A, nonlinear dependence characterizes the process of drug activation as "saturation". That is, a further increase in the feature Q_5 in the region $|Q_5| > |Q^{thr}|$ does not lead to a significant increase in radioprotective activity.

The classification rule is established in the form of a hypothesis obtained as a result of statistical analysis of experimental data. The constructed regression equation should find practical application in predictive analysis. It is well known [8] that predicting the results using regression lends itself better to meaningful interpretation than simple extrapolation of the trend (especially if the relationship is nonlinear), since this allows a fuller consideration of the nature of the phenomenon under study. However, it should be noted that with the help of the obtained regression, the evaluations of the effective indicator are carried out under averaged conditions and this should be taken into account in practical predictive studies. That is, it should be borne in mind that the regression model sets some forecasting tolerances. The forecast can be carried out using regression equations (3*b*) by substituting a numerically estimated explanatory feature into this equation. If the regression function is statistically justified [20], then forecasting has sufficient reliability. It should be emphasized that the expected values of the performance indicator are average values. Due to the diversity of phenomena and the multifaceted nature of their manifestations, the empirical values of the effective attribute are scattered around the average values. As a consequence the actual values of the performance characteristic do not have to be exactly the same as the forecast.

The validity of the proposed hypothesis can be verified by testing it on chemical compounds that were not included in the original series of compounds. An important role for models is their ability to make predictions. Let us check the ability of equation (1) to give predictive estimates. 5-Ethyltryptamine was not included in the original sample (Table 1) because the quantum mechanical calculation of the molecule electronic structure was not performed. However, it is known that the electron affinity of C_2H_5 group (Z_{sub} = 1.857 arb. units) is 1.40 eV [12], which is near the discontinuity. This affinity value falls within the confidence interval (4). Using equation (2) in Fig, 2A for this value of the energy of affinity, the value of the attribute Q_{52} = -0.27 arb. units was calculated. For comparison, we present the value Q_{51} = - 0.11 arb. units obtained from equation (1) (Fig. 2B). According to the data in table 2, the radioprotective activity of 5-ethyltryptamine should be approximately the same as that of mexamine. Indeed, from equation (1) in Fig, 3B it follows *Act*1 = 55.0%. To determine the regression (1), a sample

was used that contains only drugs with $Act1_{exp}$ activity ≥ 35%. The experiment [2] indicates that the prophylactic effect of 5-ethyltryptamine (Act_{exp} = 60%; dose 58mg/kg) does not differ from the effect of mexamine. We will also estimation the survival rate when using 5-acetyloxytryptamine. For this drug, quantum mechanical calculations of the electronic structure of the molecule were also not performed. However, the affinity energy of the substituent CH_3COO is known: A = 4.1 eV [12]. From equation (2) in Fig. 2A, we obtain the following very approximate (since the value of A is outside the confidence interval) estimate of the factor Q_{52} = -0.53 arb. units. Using regression (1) (Fig. 3B) we obtain the following survival rate estimate: *Act*1 = 71.7%. For the substituent CH_3COO the factor Z_{sub} = 3.29 arb. units Consequently, from equation (2) in Fig. 2B we obtain the estimate Q_{52} = -0.31 arb. units. In this case, the bioactivity *Act*1 = 57.6%. The observed bioactivity is 50% when using a dose of 31mg/kg [2].

A similar situation, apparently, takes place for 5-acetyltryptamine ($R_5 = H_3CCO$; Z_{sub} = 2.833 arb. units). For this substituent, the experimental value of the electron affinity has not been found in the literature. However, it is known from indirect data that the electron affinity of a group of H_3CCO atoms is not lower than 1.0 eV. From equation (1) in Fig. 2A, this affinity value corresponds to the value Q_{51} = - 0.096 < Q_5^{thr} ≈ - 0.09 arb. units (Fig. 3B). That is, in this case, the expected radioprotective effect of 5-acetyltryptamine ($R_5 = H_3CCO$) should also be approximately at the level of mexamine (from equation (1) in Fig. 3B we obtain *Act*1 = 45.4%). Using equation (2) in Fig. 2B, we obtain the following value for the factor Q_{52} = -0.29 arb. units. This value of the factor Q_{52} corresponds to a bioactivity of 56.4% (equation (1) in Fig. 3B.). The observed bioactivity value is 54.3% (the dose used is 63mg/kg [2]).

Efficiency estimates can also be obtained for 5-propoxytryptamine, for which no quantum mechanical calculations have been performed. The experimental value of the affinity energy of the atomic group $R_5 = H_7C_3O$ (Z_{sub} = 2.273 arb. units) is known and is equal to A = 0.67 eV [12]. This value of the affinity energy corresponds to the estimate of the molecular parameter Q_5 = - 0.18 arb. units (Q_{51} = - 0.13 arb. units; equation (1) in Fig. 2B). Using equation (2) in Fig. 3B, we obtain the following estimate of the radioprotective effect of 5-propoxytryptamine: *Act*2 = 9.3% (for the factor Z_{sub}: *Act*2 = 8.1%), which is close to the observed value of drug efficacy (Act_{exp} = 13.3%; dose 50mg/kg [2]). It should be noted that the value Q_5 = -0.09 arb. units is located near the threshold value Q_5^{thr} ≈ -0.09 arb. units. Note that the value Z_{sub} = 2.273 arb. units is also near the boundary value. Beyond the threshold there is a sharp jump in bioactivity (Fig. 3B). When constructing regression 2 (Fig. 3B), a sample containing only drugs with Act_{exp} ≤ 10% activity was used. There is some uncertainty when assessing the bioactivity of drugs near the threshold. For example, if we use non-linear regression (Act_{nonl}, Fig. 3B), combining the entire sample (Fig. 3A), the following estimate of the radioprotective activity of 5-propoxytryptamine was obtained: *Act*3 = 35.8%.

Let us estimate the radioprotective activity, for example, of 5-acetyloxytryptamine ($R_5 = CH_3COO$). The drug was not included in the original sample because it was tested at a dose of 31 mg/kg (a dose significantly lower than the equimolar dose of 50mg/kg tryptamine). We will also estimate the bioactivity of 5-ethoxytryptamine ($R_5 = H_5C_2O$; Act_{exp} = 16%; dose 75 mg/kg) [2]. For the first compound, the value of the molecular feature calculated by the quantum mechanical method is Q_5 = -0.5387 arb. units. For the group of atoms CH_3COO the value of the affinity energy to the electron A = 4.1 eV [12]. Using equation (1), an independent estimate of the feature Q_5 can be specified. The following value was obtained Q_5 = - 0.564 arb. units, which is very close to the value determined by the quantum mechanical method. From the linear equation (1) in Fig. 3B, it follows that the theoretical estimate of bioactivity is 41%. This value of anti-radiation protection is close to the known experimental value Act_{exp} = 50% [2]. For comparison, we also give an estimate of bioactivity using a non-linear equation (Fig. 3B). The result was equal to 58.8%. For 5-ethoxytryptamine, quantum mechanical calculations gave the value of the feature Q_5 = -0.1812 arb. units. For this substituent, there is an approximate estimate of the electron affinity energy A ≈ 1.60 eV [12]. From equation (1) (Fig. 2A) we obtain the following estimate of the value of the attribute Q_5 = - 0.1938 arb. units, which is close to the value calculated by the quantum mechanical method. Using the value Q_5 = -0.1812 arb. units in equation (1) (Fig. 3B) we obtain the bioactivity estimate: *Act*1 = 49.3%. Consequently, as expected, a decrease (in absolute value) in the negative sign Q_5 (Fig. 3B) reduces the radioprotective activity of the drug. The calculated value of the radioprotective efficacy of the drug is noticeably higher than the observed one (Act_{exp} = 16%, dose 75 mg/kg). Perhaps this is due to the proximity of the Q_5 feature to its threshold value or

is associated with the steric properties of the substituent (see below). For example, if we use equation (2) (slightly to the right of the threshold; see Fig. 3B), then we obtain the bioactivity estimate *Act*2 = 9.3% for 5-ethoxytryptamine.

It can also be noted that an increase in the number of carbon and hydrogen atoms of the substituent in the fifth position of the benzene ring leads to a decrease in the radioprotective effect in comparison with 5-methoxytryptamine (Q_5 = - 0.2767 arb.units, dose 75 mg/kg, Act_{exp} = 69.3%; linear the model gives the following estimate of activity: *Act*1 = 53.7% (Fig. 3B); the nonlinear model leads to the value: Act_{nonl} = 65.7%). An increase in the number of atomic CH_2 groups in the molecule is accompanied, on the one hand, by a 1.5-fold decrease in the indicator Q_5 (in absolute value) and, on the other hand, by an increase in the hydrophobic properties of the molecule.

Let us now analyze the sample (Table 1), which contains only chemical compounds (Nos. 1 - 12) with a substituent in the fifth position (R_5). The sample also included the following drugs: tryptamine (in this case, it can be formally assumed that $R_5 \equiv H$) and 5-methoxy-7-chlorotryptamine. Let us find out what property of the R_5 substituent can have a significant effect on the bioactivity of the molecule. Let us check the existence relationship between the volume size (*MR* – molar volume) of the substituent R_5 and the bioactivity of molecules. Previously, the following significant regression was obtained, taking into account the relationship between the radioprotective properties of drugs and the value of the electronic factor Q_5:

$Act/100 = a_0 + a_1 \cdot Q_5$, N = 12, m_1 = 1; R = - 0.77 ± 0.13, $|R^*| = 0.79 > R_{0.05}^{cr}(N - 2) = 0.576$; criterion for the significance of the correlation coefficient based on the Fisher normalizing z-transform (taking into account Hotelling's corrections): $u_H = 0.988 > u_{0.05}(N) = z_{0.975} \cdot (N - 1)^{-0.5} = 0.591$; $RMSE$ = 0.155; minimum sample size sufficient for the reliability of the correlation coefficient: $N_{0.05}^{min} = 6$; a_0 = 0.17 ± 0.09, a_1 = – 1.01 ± 0.26, $|t(a_1)| = 3.8 > t_{0.05}^{cr}(N - 2) = 2.228 > t(a_0) = 1.93$; unexplained regression residuals (disturbing variable [8]) are normally distributed: Wilk-Shapiro test: $W = 0.949 > W_{0.05}^{cr}(N) = 0.859$; $F = 14.6 > F_{0.05}^{cr}(f_1 = 1; f_2 = 10) = 4.28$; straightforwardness factor: $K = 2.12 < K^{thr} = 3.00$; $\Sigma_1 = 0.2392$.

(19)

Note that a twofold decrease in the sample size does not lead to a significant difference between regression (19) and regression (9). That is, we can assume the presence of consistency properties of the estimates of the regression parameters. Cause-and-effect relationships (9), characterizing the putative mechanism of bioactivity of chemical compounds, hold out with a significant change in the sample size. Statistics of the sets Q_5 and $Act_{exp}/100\% \equiv Act$ will be as follows:

Q_5: N = 12, Q_5^{av} = - 0.28 ± 0.05; 95% confidence interval: (-0.40, -0.17); Q_5^{min} = - 0.504, Q_5^{max} = - 0.0548, S_Q = 0.177, $\tau^{min} = 1.24 < \tau^{max} = 1.30 < \tau_{0.05}^{cr,2}(N) = 2.387 < \tau_{0.05}^{cr,1}(N) = 2.523$; Wilk-Shapiro normality test: $W = 0.853 \approx W_{0.05}^{cr}(N) = 0.859$; David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(N) = 2.800 > U = [(Q_5^{max} - Q_5^{min})/S_Q] = 2.54 < U2_{0.05}^{cr}(N = 3.910$;

Act: N = 12, Act^{av} = 0.46 ± 0.07; 95% confidence interval: 0.31 – 0.61; Act^{min} = 0, Act^{max} = 0.693, S_{Act} = 0.238, $\tau^{max} = 0.97 < \tau^{min} = 1.98 < \tau_{0.05}^{cr,2}(N) = 2.387 < \tau_{0.05}^{cr,1}(N) = 2.523$; Wilk-Shapiro normality test: $W = 0.860 > W_{0.05}^{cr}(N) = 0.859$; David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(N) = 2.800 < U = [(Act^{max} - Act^{min})/S_{Act}] = 2.91 < U2_{0.05}^{cr}(N = 3.910$.

(20)

In order to clarify the significance of the influence of the volumetric size of a substituent on the bioactivity of molecules and its independence from the factor attribute Q_5, the residuals of $\delta Act/100$ regression (19) were analyzed. The residuals have a distribution close to the normal distribution: $W = 0.949 > W_{0.05}^{cr}(N) = 0.859$.

Since the results obtained above indicate the importance of taking into account the electronic properties of the substituent, the problem naturally arises of checking the relationship between the geometric size and the variability of the effective feature. As a characteristic of the steric size of a substituent, we will use its molar refraction (*MR*), which has the dimension of the volume of the atomic group (cm^3/mol). Let's first find out what possible functional relationship is linear or nonlinear between the *MR* variable and the residuals $\delta Act/100$ of the regression (19). For this we use method (14). In accordance with the Sturges rule (14), we divide the sample ranked by the *MR* attribute (i.e., ordered by size) into $n = 1 + 3.322 \cdot \lg(12) \approx 4$ approximately equal groups; i, j = 1,2, ...,4. We obtain the following comparative values of the parameters $b_{MR/A,i\text{-}j}$ for groups i and j of the method (14):

$b_{MR/A,1\text{-}2} = (\delta Act_1^{av} - \delta Act_2^{av})/(MR_1^{av} - MR_2^{av}) =$

0.00118,

$b_{MR/A,1\text{-}3} = (\delta Act_1^{av} - \delta Act_3^{av})/(MR_1^{av} - MR_3^{av}) =$

0.0073,

$b_{MR/A,2\text{-}3} = (\delta Act_2^{av} - \delta Act_3^{av})/(MR_2^{av} - MR_3^{av}) =$

-0.00074,

$b_{MR/A,1\text{-}4} = (\delta Act_1^{av} - \delta Act_4^{av})/(MR_1^{av} - MR_4^{av}) =$

0.0086. (21)

Here i and j are group numbers. The values of the parameters $b_{MR/A,i\text{-}j}$ must be very close to each other for a straightforward relationship of features. However, it follows from relations (21) that these parameters differ significantly not only in magnitude, but also in sign. Therefore, the relationship between the feature MR and the value of the $\delta Act/100$ residues should be non-linear. We can test quadratic two-factor regression:

$$\delta Act/100 = b_0 + b_1 \cdot MR + b_2 \cdot MR^2,$$

$N = 12$, $m = 2$; multiple correlation coefficient: $R = 0.81 > R_{0.05}^{cr}(v = N - m - 1; f = m) = 0.697$ [26], $R^2 = 0.66$; $RMSE = 0.097$, $b_0 = -0.05 \pm 0.07$, $b_1 = 0.026 \pm 0.013$, $b_2 = -1.24 \cdot 10^{-3} \pm 4 \cdot 10^{-4}$, $|t(b_2)| = 2.86 > t_{0.05}^{cr}(f = 9) = 2.262 > t(b_1) = 2.006$; $F = 8.33 > F_{0.05}^{cr}(f_1 = 2; f_2 = 9) = 4.26$; $\Sigma_2 = 0.0842$. (22)

In statistics (22), the $R_{0.05}^{cr}(v = 9; m = 2)$ test (here $v = N - m - 1$, m is the number of explanatory variables [26]) indicates the critical value of the sample multiple correlation coefficient. The statistical significance of multiple regression can be estimated using the F-statistic. Since the empirical value $F > F_{0.05}^{cr}(v = 2; m = 9)$, therefore, the relationship between a result and at least one explanatory variable can be considered statistically significant. Therefore, the independent variable MR^2 is significantly associated with variation in bioactivity, which was not accounted for when using the explanatory variable Q_5. In accordance with the results (22), only the regression coefficient b_2 will be significant, which determines the nonlinearity of the relationship. It also follows from inequalities (22) that at the 95% confidence level the t-value of the regression coefficient b_1 is below the critical level, that is, its difference from zero at the significance level $\alpha = 0.05$ can be considered random. The sample statistics of the $MR(R_5)$ and $MR^2(R_5)$ sets are as follows:

MR: $N = 12$, $MR^{av} = 9.94 \pm 2.50$; 95% confidence interval: 4.44 - 15.44; $MR^{min} = 0.92$, $MR^{max} = 27.68$, $S_{MR} = 8.66$, $\tau^{min} = 1.04 < \tau^{max} = 2.05 < \tau_{0.05}^{cr,2}(N) = 2.387 < \tau_{0.05}^{cr,1}(N = 2.523$; David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(N) = 2.800 < U = [(MR^{max} - MR^{min})/S_{MR}] = 3.09 < U2_{0.05}^{cr}(N) = 3.910$;

MR^2: $N = 12$, $MR^{2,av} = 167.5 \pm 74.5$; 95% confidence interval: (3.41 - 331.6); $MR^{2,min} = 0.846$, $MR^{2,max} = 766.2$, $S_{MR2} = 258.3$, $\tau^{min} = 0.65 < \tau^{max} = 2.32 < \tau_{0.05}^{cr,2}(N) = 2.387 < \tau_{0.05}^{cr,1}(N = 2.523$; David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(N) = 2.800 < U = [(MR^{2,max} - MR^{2,min})/S_{MR2}] = 2.963 < U2_{0.05}^{cr}(N) = 3.910$. (23)

Let us apply the quadratic approximation (22) to the observed bioactivity (Table 1):

$$Act4/100 = a_0 + a_1 \cdot MR + a_2 \cdot MR^2,$$

$N = 12$, $m = 2$; multiple correlation coefficient: $R = 0.86 > R_{0.05}^{cr}(v = N - m - 1; f = m) = 0.697$, $R^2 = 0.74$; $RMSE = 0.133$, $a_0 = 0.392 \pm 0.093$, $a_1 = 0.039 \pm 0.018$, $a_2 = -1.96 \cdot 10^{-3} \pm 6 \cdot 10^{-4}$, $|t(a_2)| = 3.30 > t_{0.05}^{cr}(f = 9) = 2.262 > t(a_1) = 2.006$; $F = 12.26 > F_{0.05}^{cr}(f_1 = 2; f_2 = 9) = 4.26$; $\Sigma_4 = 0.1583$, AIC = -3.9948, SC = -3.7069, SS = 0.0398. (24)

Regression (24) does not contradict regression (22). Features MR and MR^2 are closely related. The pair correlation coefficient is 0.97. This leads to collinear explanatory variables.

One of the known ways to reduce the interconnectedness of features is a linear transformation of variables. Let's introduce new variables: $\Delta 1 = MR - MR^{av}$ и $\Delta 2 = (MR - MR^{av})^2$. The regression equation can now be written as follows:

$$Act5/100 = b_0 + b_1 \cdot \Delta 1 + b_2 \cdot \Delta 2, \quad (25)$$

$N = 12$, $m = 2$; multiple correlation coefficient: $R_{41} = 0.86 > R_{0.05}^{cr}(v = N - m - 1; f = m) = 0.697$, $R_{41}^2 = 0.732$, $R_{41}^{*2} = 0.68$; $RMSE = 0.133$, $b_0 = 0.59 \pm 0.06$, $b_1 = 0.0001 \pm 0.0077$, $b_2 = -1.19 \cdot 10^{-3} \pm 6.62 \cdot 10^{-4}$, $t(b_0) = 9.41 > |t(b_2)| = 2.91 > t_{0.05}^{cr}(f = N - m - 1) = 2.262 > t(b_1) = 0.07$; $F = 12.26 > F_{0.05}^{cr}(f_1 = 2; f_2 = 9) = 4.26$; $\Sigma_1 = 0.1582$; $AIC_1 = -3,9948$, $SC_1 = -3.7076$, $SS_1 = 0.0398$.

The correlation coefficient between the variables $\Delta 1$ and $\Delta 2$ decreases: $r_{\Delta 1,\Delta 2} = 0.755$. Statistics (24) and (25) give the Akaike information criterion [27] of the relative quality of a linear statistical model for a given data set and is determined as follows:

$$AIC = 2m/N + \ln(\Sigma/N). \quad (26)$$

Here m is the number of explanatory variables in the regression; Σ is the sum of the squares of the

regression residuals; N is the number of observations. The regression quality test (26) establishes a trade-off between the residual sum of squares and the number of explanatory variables. The regression residuals are assumed to be normally distributed. In this case, the Wilk-Shapiro test of normality will be as follows: $W = 0.907 > W_{0.05}^{cr}(N) = 0.859$. The test formula (26) usually contains the constant $1 + \ln(2\pi)$. This constant is not taken into account here, since it is not essential for comparative tests.

The Akaike test quantifies the relative amount of information that is lost when building a statistical model. The less information is lost (that is, the smaller the AIC value (26)), the higher the quality of the model. When comparing statistical models, preference is given to the model for which the AIC test is the smallest. That is, the model minimizes the loss of information. The test is useful only when comparing statistical models, and the size of the compared samples N must be the same. The absolute value of the test is not very informative. It is important to note that the compared linear models should have samples of equal size, and the regression residuals have a distribution close to the normal distribution. Recently, the Schwarz criterion is also often used [28]:

$$SC = (m + 1)\ln(N)/N + \ln(\Sigma/N). \quad (27)$$

The estimate obtained using this indicator is considered consistent. For a relative estimation of the quality of the model, you can also use the relation: $SS = \Sigma^{1/2}/(N - m)$. This ratio is always associated with the values of the Akaike and Schwartz tests.

Let's go back to regression (25). A linear transformation reduces the significance of the explanatory variable Δ1 to almost zero: $t(b_1) = 0.07$ (compare with (22)). The statistics of the sets Δ1 and Δ2 will be as follows:

Δ1: $N = 12$, $\Delta1^{av} = 0.00 \pm 2.50$; 95% confidence interval: (-5.50, 5.49); $\Delta1^{min} = -9.02$, $\Delta1^{max} = 17.74$, $S_{\Delta1} = 8.66$, $\tau^{min} = 1.04 < \tau^{max} = 2.05 < \tau_{0.05}^{cr,2}(N) = 2.387 < \tau_{0.05}^{cr,1}(N = 2.523$; David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(N) = 2.800 < U = [(\Delta1^{max} - \Delta1^{min})/S_{\Delta}] = 3.09 < U2_{0.05}^{cr}(N) = 3.910$;

Δ2: $N = 12$, $\Delta2^{av} = 66.70 \pm 29.55$; 95% confidence interval: 3.67-133.7; $\Delta2^{min} = 1.225$, $\Delta2^{max} = 314.76$, $S_{\Delta2} = 102.36$, $\tau^{min} = 0.66 < \tau^{max} = 2.05 < \tau_{0.05}^{cr,2}(N) = 2.387 < \tau_{0.05}^{cr,1}(N = 2.523$; David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(N) = 2.800 < U = [(\Delta2^{max} - \Delta2^{min})/S_{\Delta}] = 3.00 < U2_{0.05}^{cr}(N) = 3.910$. (28)

Standardized regression coefficients:

$$b_1^* = b_1 \cdot S_{\Delta1}/S_{Act} = -0.021,$$
$$b_2^* = b_2 \cdot S_{\Delta2}/S_{Act} = -0.839. \quad (29)$$

Using the standardized coefficients (29), it is possible to obtain an estimate of the values of the relative contributions of the explanatory variables Δ1 and Δ2 to the variability of the bioactivity:

$$R_{пр}^2 = b_1^* \cdot r_{\Delta1,Act} + b_2^* \cdot r_{\Delta2,Act} =$$
$$0.015 + 0.717 = 0.732. \quad (30)$$

Here $r_{\Delta1,Act} = -0.69$ and $r_{\Delta2,Act} = -0.86$ are the correlation coefficients of the explanatory variables Δ1 and Δ2 with the observed bioactivity. The approximate value of the coefficient of determination (30) practically coincides with the coefficient of determination (25). From the relation (30) it follows that the dominant contribution to the variability of bioactivity is made by the variable Δ2.

The use of linearly transformed explanatory variables clearly indicates the insignificance of the Δ1 variable in the regression equation. The $t(b_1)$ value is significantly lower than the table value at $\alpha = 0.05$. Therefore, at the 95% confidence level, the variable Δ1 is statistically insignificant. The statistical insignificance of the Δ1 variable can also be established as follows. To do this, it is necessary to analyze a regression in which only one explanatory variable Δ2 is taken into account. Next, the regression residuals $Act(\Delta2)$ are calculated, for which their relationship with the explanatory variable Δ1 is checked. It turned out that the correlation coefficient between the sets of residuals and the value of Δ1 is insignificant: $r = 0.017$. Therefore, the explanatory variable Δ1 in regression (25) can be ignored.

As shown above, the bioactivity of drugs is associated with the value of the complex electronic factor Q_5. Therefore, let us find out whether it is possible to supplement regression (25) with one more independent explanatory variable, namely Q_5. Let's first check whether there is a relationship between the regression residuals $Act(\Delta2)$ and the population Q_5. The following statistics were obtained: $|R^*| = 0.76 > R_{0.05}^{cr}(N - 2) = 0.576$, $F = 13.33 > F_{0.05}^{cr}(f_1=1; f_2=10) = 4.96$. Thus, the unexplained regression residuals (25) are significantly correlated with the value of the Q_5

variable and, therefore, this variable can be included in the regression.

Let us compose a two-factor regression, which, as explanatory variables, contains the molecular electronic feature Q_5 and the geometric factor of the substituent, that is, its molecular refraction *MR*. Preliminarily check for collinearity between the explanatory variables Q_5 and Δ2. The correlation coefficient between the variables Q_5 and Δ2 turned out to be equal to $r_{1,2} = 0.35 < R_{0.05}^{cr}(N - 2) = 0.576$, which is less than the generally accepted maximum allowable value of 0.8 [8] for the explanatory variables. The statistics of the ratio $F = 1.42 < F_{0.05}^{cr}$ ($f_1 = 1$; $f_2 = 10$) = 4.96 also indicates the insignificance of the interrelation of features.

Now we will compose a regression that takes into account the simultaneous combined effect of the explanatory variables Q_5, Δ1, and Δ2 on bioactivity. These variables characterize different aspects (i.e., the electronic properties of the molecule and the geometric dimensions of R_5) of the independent influence of the substituent in the fifth position of the indole ring on the variability of the bioactivity of the molecule:

$$Act6/100 = b_0 + b_1 \cdot Q_5 + b_2 \cdot \Delta 1 + b_3 \cdot \Delta 2, \quad (31)$$

$N = 12$, $m = 3$, $R_6 = 0.96 > R_{0.05}^{cr}(v = N - m - 1; f = m) = 0.777$, $R_6^2 = 0.923$, $R_6^{*2} = 0.894$; *RMSE* = 0.075, $b_0 = 0.38 \pm 0.06$, $b_1 = -0.65 \pm 0.14$, $b_2 = 2.0 \cdot 10^{-3} \pm 4.4 \cdot 10^{-3}$, $b_3 = -1.6 \cdot 10^{-3} \pm 3.8 \cdot 10^{-4}$, $t(b_0) = 6.65 > |t(b_3)| = 4.22 > |t(b_1)| = 4.48 > t_{0.05}^{cr}(f = 8) = 2.306 > |t(b_2)| = 0.50$; $F = 32.19 > F_{0.05}^{cr}(f_1 = 3; f_2 = 8) = 4.07$; $b_1^* = b_1 \cdot S_Q / S_{Act} = -0.49$, $b_2^* = b_2 \cdot S_{\Delta 1}/S_{Act} = 0.08$, $b_3^* = b_3 \cdot S_{\Delta 2}/S_{Act} = -0.70$; $R_{appr1}^2 = b_1^* \cdot r_{Q,Act} + b_2^* \cdot r_{\Delta 1,Act} + b_3^* \cdot r_{\Delta 2,Act} = 0.377 - 0.055 + 0.602 = 0.924$; $\Sigma_6 = 0.0462$; $AIC_6 = -5.0597$, $SC_6 = -4.7314$, $SS_6 = 0.0239$.

The addition of the independent explanatory variable Q_5 to the regression (25) significantly reduces the information criteria. That is, the loss of information decreases. Comparing the Akaike (AIC), Schwarz (SC) tests, and SS ratios for regressions (25) and (31), it is easy to see that the relative quality of regression (31) for all three tests is higher than the quality of regression (25).

The standardized coefficients b_i^* characterize the relative strength of the influence of each explanatory variable on the bioactivity. As well as for regression (25), the explanatory variable Δ1 is statistically insignificant. In statistics (31), the adjusted coefficient of determination is given $R_6^{*2} = 1 - (1 - R_6^2)(N - 1)/(N - m - 1)$ [8], taking into account the change in the number of explanatory variables. Applying a correction for the number of degrees of freedom gives the unbiased estimate of determination coefficient. This transformation is essential for small sample sizes ($N < 20$); *m* is the number of explanatory variables. The preference is given to the regression for which the corrected coefficient of determination is greater. The transition to standardized regression coefficients b_i^* makes it possible to compare the relative proportional effect of explanatory variables on the variability of bioactivity. The correlation coefficients are $r_{Q,Act} = -0.77$, $r_{\Delta 1,Act} = -0.69$, $r_{\Delta 2,Act} = -0.86$, respectively. Since $|t(b_2)|$ is significantly less than the table value, then the regression coefficient b_2 differs insignificantly from zero, at a confidence level of 95%. Therefore, the explanatory variable Δ1 in equation (31) can be ignored. This is also indicated by the values of the relative contributions to the approximate value of the coefficient of determination R_{appr1}^2. The value $R_{appr1}^2 = 0.924$ practically coincides with the value $R_6^2 = 0.923$ (31). Since the regression residuals (31) are normally distributed (Wilk-Shapiro test: $W = 0.862 > W_{0.05}^{cr}(N) = 0.859$), then to quantify the collinearity of the remaining explanatory variables Q_5 and Δ2, we can use the Farrar-Glauber relation [8,17] (pair correlation coefficient $r_{1,2} = 0.35$):

$$\chi^2 = -(N - 1 - (2m + 5)/6) \cdot \ln\left(\det \begin{vmatrix} r_{1,1} & r_{1,2} \\ r_{2,1} & r_{2,2} \end{vmatrix}\right) = 1.24 < \chi_{0.05}^{2,cr}(f = 1) = 3.841, \quad (32)$$

and also *t*-statistics (2), which is valid for $N \geq 10$ [16]:

$$t = r_{1,2} \cdot (N - m)^{0.5}/(1 - r_{1,2}^2)^{0.5} = 1.18 < t_{0.05}^{cr}(f = N - m) = 2.228. \quad (33)$$

Here the sample size is $N = 12$; $m = 2$ is the number of explanatory variables. It follows from inequalities (32) and (33) that at the 95% confidence level, collinearity between the explanatory variables can be neglected. Thus, it can be agreed that the explanatory variables Q_5 and Δ2 simultaneously and independently affect the variability of the bioactivity.

According to the inequality $t_{0.05}^{cr} > |t(b_2)| = 0.50$ instead of the regression (31) the following regression can be used:

$$Act6_1/100 = b_0 + b_1 \cdot Q_5 + b_2 \cdot \Delta 2,$$

$N = 12$, $m_2 = 2$, $R_{61} = 0.96 > R_{0.05}^{cr}(v = N - m - 1; f = m) = 0.697$, $R_{61}^2 = 0.921$, $R_{61}^{*2} = 0.903$; *RMSE* = 0.072, $b_0 = 0.37 \pm 0.05$, $b_1 = -0.64 \pm 0.14$, $b_2 = -1.4 \cdot 10^{-3} \pm 2.4 \cdot 10^{-4}$, $t(b_0) = 7.11 > |t(b_2)| = 6.11 >$

$|t(b_1)| = 4.65 > t_{0.05}^{cr}(f = 9) = 2.262$; $F = 52.47 > F_{0.05}^{cr}(f_1 = 2; f_2 = 9) = 4.26$; $\Sigma_{61} = 0.0466$; $AIC_{61} = -5.2177$, $SC_{61} = -4.9298$, $SS_{61} = 0.0216$. (34)

Regression residuals (34) are normally distributed. The Wilk-Shapiro test is as follows: $W = 0.907 > W_{0.05}^{cr}(N) = 0.859$. The regression quality (34) is higher than the regression quality (31) for all three AIC, SC and SS tests. Therefore, excluding the variable Δ1 from the regression does not lead to a comparative worsening of the regression. Regression (34) cannot explain only 9.7% of the total variance. The quantity of the uncertainty coefficient of 0.097 may be due to the influence of some other unaccounted for molecular characteristics. The fact that the variable Δ1 is not included in the regression equation does not mean that there is no dependence of the response on the linear size of *MR*. The linear variable holds because it is part of the explanatory variable $\Delta 2 = (MR - MR^{av})^2$.

The regression coefficients (34) are dimensions. It is convenient to go to standardized coefficients in order to find out the relative role of each variable of the regression equation in the variability of the bioresponse:

$$b_1^* = b_1 \cdot S_Q / S_{Act} = -0.49,\ b_2^* = b_2 \cdot S_{\Delta 2} / S_{Act} = -0.64. \quad (35)$$

The contributions sharing to the determinative coefficient of significant explanatory variables remain relevant, as they are for regression (31). The approximate coefficient of determination practically coincides with the coefficient $R_6^2 = 0.921$ (34):

$$R_{appr}^2 = b_1^* \cdot r_{Q,Act} + b_2^* \cdot r_{\Delta 2,Act} =$$

$$0.374 + 0.546 = 0.920. \quad (36)$$

Here $r_{Q,Act} = -0.77$ and $r_{\Delta 2,Act} = -0.855$ are the correlation coefficients between the independent explanatory variables Q_5, Δ2 and the observed bioactivity (Act), respectively. Thus, variable Δ2 makes the largest contribution of 54.6% in explaining the variability of bioactivity. This represents more than 62% of the explained response variability, whereas for variable Q_5 this contribution is 37.4% only. Unexplained (or random) remains 8% of the variance of the regression (34).

Further information on how significant the contribution of Δ2 is in explaining variations in bioactivity can be obtained by comparing the determination coefficients of regressions (19) and (34). Let us use the relation [8] for this:

$$F = (R_{61}^2 - R^2)(N - m_2 - 1)/(m_2 - m_1)/(1 - R_{61}^2) =$$

$$37.7 > F_{0.05}^{cr}(f_1 = 1; f_2 = 9) = 4.26, \quad (37)$$

which has an *F*-distribution with degrees of freedom: $f_1 = m_2 - m_1$, $f_2 = N - m_2 - m_1$. Since $F > F^{cr}$, the null hypothesis is rejected at a significance level of $\alpha = 0.05$. Consequently, it can be agreed that the additional explanatory variable Δ2 contributes a significant proportion in explaining the variation in bioactivity, and the joint effect of the explanatory variables Q_5 and Δ2 on the variation in bioactivity is significant.

Since $R_6 > R^{cr}$ the null hypothesis of a zero multiple correlation coefficient can be rejected at the significance level $\alpha = 0.05$. In the specialist literature, it is generally accepted that a model is acceptable if the coefficient of determination is greater than 0.5. From the regression analysis (34) it also follows, that significant decrease of radioprotective action of preparations occurs at increase of both volumetric size of substituent $MR(R_5)$, and complex electronic feature Q_5 (direction of increase from negative value to positive value).

It was established above that equations (24) and (25) are identical, i.e. using *MR* and Δ as explanatory variables is equivalent. It is therefore of interest to construct a regression equation that uses the explanatory variables Q_5, *MR* and MR^2 simultaneously (compare with regression (31). This regression is of interest because it allows an approximate estimate to be made of the optimal value of the *MR* factor of the substitutes. The three-factor regression would have the following form:

$$Act7/100 = b_0 + b_1 \cdot Q_5 + b_2 \cdot MR + b_2 \cdot MR^2,$$

$N = 12$, $m = 3$, $R_7 = 0.96 > R_{0.95}^{cr}(\nu = 8; m = 3) = 0.777$, $R_7^2 = 0.922$, $R_7^{*2} = 0.892$; $RMSE = 0.076$, $b_0 = 0.22 \pm 0.07$, $b_1 = -0.64 \pm 0.15$, $b_2 = 0.031 \pm 0.010$, $b_3 = -1.5 \cdot 10^{-3} \pm 3.6 \cdot 10^{-4}$, $|t(b_1)| = 4.40 > |t(b_3)| = 4.23 > t(b_0) = 3.32 > t(b_2) = 2.99 > t_{0.05}^{cr}(f = 8) = 2.306$; $F = 31.48 > F_{0.05}^{cr}(f_1 = 3; f_2 = 8) = 4.07$; $\Sigma_{71} = 0.0462$; $AIC_{71} = -5.0596$, $SC_{71} = -4.732$, $SS_{71} = 0.0358$. (38)

The information quality criteria are inferior to the regression (34). However, all regression coefficients (38) are significantly different from zero at the 95% confidence level when taken together simultaneously. This result gives an approximate estimate from regression (38) of the optimum value of the explanatory feature *MR* for a sample that contains only chemical compounds with substituent

R_5 in the fifth position. After differentiating the *Act*7/100 regression with respect to the *MR* variable, we obtain the optimal value of the volumetric size of the substituent:

$$MR(R_5)_{oopt} = - b_2/(2 \cdot b_3) = 10.3 \text{ cm}^3/\text{mol}. \quad (39)$$

From the regression (31) we obtain a close to optimum value of $MR(R_5)_{opt} = 10.2$ cm^3/mol.

Using regression equations (26) and (23), let us evaluate the bioactivity of the drug 5-benzyloxytryptamine ($R_5 = C_6H_5CH_2O$; dose 50mg/kg). This drug was not included in the original sample. For this substituent, the value of the characteristic $MR(R_5) = 32.19$cm^3/mol is known [29,30]. An approximate estimate of the affinity energy can also be given: A ≈ 1.4 eV [12]. From equation (2) in Fig. 2A, the following estimate of the magnitude of the complex feature was obtained: Q_{52}(A) = - 0.28 arb. units. For the substituent $R_5 = C_6H_5CH_2O$, the electronic factor $Z_{sub} = 2.733$ arb. units. For this factor Z_{sub} we obtain from equation (2) in Fig. 2B the value $Q_{52}(Z_{sub}) = 0.286$ arb. units, which is very close to Q_{52}(A) = -0.28 arb. units. Using the values of the signs *MR* and Q_{52}(A) from equations (23) or (26), we obtain that 5-benzyloxytryptamine should not have a radioprotective effect. This result does not contradict the observations [2,6]. From equations (34) and (38), the bioactivity is negative. This is due to the fact that the values of the explanatory variables *MR* and Q_5 go beyond the confidence intervals (23) and (20).

Similarly, an estimate of the bioactivity of 5-acetyloxytryptamine can be obtained ($R_5 = CH_3COO$; $Act_{exp} = 50\%$, A = 3.36 ± 0.05 eV [12], $MR(R_5) = 12.87$ cm^3/mol). The original sample did not include 5-acetyloxytryptamine because the experiment was performed using a dose of 31 mg/kg [2]. Calculated according to equation (2) in Fig. 2A, the value of the feature Q_{52}(A) = -0.456 arb. units. Using equation (23) or (26) the following values for the radioprotective effectiveness of 5-acetyloxytryptamine were obtained: 65% (66.2%), which are comparable to the observed radioprotective effect of $Act_{exp} = 50\%$. The first value is obtained from equation (34), and the second value, in parentheses, is obtained from equation (38). In assessing bioactivity we will follow this sequence in the future.

Bioactivity was also assessed for 5-ethyltryptamimine. ($R_5 = C_2H_5$; $Act_{exp} = 65\%$, dose 58 mg/kg; $Z_{sub} = 1.86$ arb. units, A = (0.9 - 1.4) eV [12], $MR(R_5) = 10.27$ cm^3/mol). In accordance with Figs. 2A and 2B, for the affinity interval A, the complex electronic parameter Q_5 is Q_{51}(A) = -0.080 arb. units, Q_{52}(A) = -0.27 arb. units, respectively. According to model (34), the bioactivity estimate is in the range (42 - 54)%. For the factor $Z_{sub} = 1.8572$ arb. units, the parameter $Q_{51}(Z) = -0.112$ arb. units was obtained, which does not contradict the interval obtained using the electron affinity energy.

Above, an assessment of the effectiveness of 5-ethoxytryptamine has already been carried out (*Act*1 = 56%; equation (1) of Fig. 3B; $Act_{exp} = 16\%$, dose 75 mg/kg). Now let us evaluate the bioactivity taking into account the volumetric size of the substituent $R_5 \equiv H_5C_2O$ (calculated quantum value $Q_5 = -0.1812$ arb. units; $MR(R_5) = 12.47$ cm^3/mol). Using two-factor regression (34) and three-factor regression (38), we obtain the following estimates of the effectiveness of the drug 49.4% (47.6%). That is, taking into account the volumetric size of the substituent significantly affects the activity of the drug, reducing the estimate (*Act*1 = 56%) of bioactivity (regression coefficient $b_2 < 0$ (34) and $b_3 < 0$ (38)). For comparison, we present the value $Q_{51}(Z) = -0.133$ arb. units (equation (1) of Fig. 2B) using the value of the electronic factor $Z_{sub} = 2.375$ arb. units.

If we accept the validity of equations (1) and (2) in Figs. 2A, and 2B, as well as equations (34) and (38), then we can assume that compounds for which substituents with a sufficiently high value of the electron affinity, for example, HCOO (A = 3.9 eV; Q_{52}(A) = -0.508 arb. units; $MR = 6.93$ cm^3/mol), NO_2 (A = 3.1 eV; Q_{52}(A) = -0.432 arb. units; $MR = 7.36$ cm^3/mol), SO_2H (A = 2.8 eV; Q_{52}(A) = -0.403 arb. units; $MR = 8.87$ cm^3/mol), CH_3COO (A = 4.1-3.36 eV; Q_{52}(A) = (-0.527, -0.456) arb. units; $MR = 12.47$ cm^3/mol). For these substituents, the estimated bioactivity will have the following approximate values (when used in doses of equimolar 50 mg/kg tryptamine): 68.2, 63.7, 62.6 and 69.8-65.2, respectively. The numerical values of the electron affinity of radicals are taken from the reference book [12]. The molar refraction values of the substituents are given in [31-32]. It should be noted that for the HCOO substitute, the value of the Q_{52} feature is outside the 95% confidence interval (20). It should be noted that for R_5 = HCOO substitute, the value of the Q_{52} feature is outside the 95% confidence interval (5.20). For $R_5 = CH_3COO$, the upper limit of Q_{52} also slightly exceeds the confidence interval.

Let us also analyze the relationship between the electronic factor *Z* and the information factor *H* [33] for a private sample containing only bioactive drugs (A > 38%; Nos. 1-9). Factors *Z* and *H* characterize

the molecule as a whole. The following significant linear relationship was found:

$H(Z)_1 = a_{01} + a_{11}\cdot Z$, $N_1 = 9$, $R_1 = 0.93 \pm 0.05$, $R_1^* = 0.94 > R_{0.05}^{cr}(N_1 - 2) = 0.666$; $RMSE(S_1) = 0.043$; sample size sufficient for the validity of the correlation coefficient: $N_{0.05}^{min} < 5$; criterion for the significance of the correlation coefficient based on the Fisher normalizing z-transform (taking into account Hotelling's corrections): $u_H = 1.567 > u_{0.05}(N_1) = z_{0.975}\cdot(N_1 - 1)^{-0.5} = 0.693$; $a_{01} = -0.56 \pm 0.32$, $a_{11} = 0.75 \pm 0.12$, $|t(a_{11})| = 6.45 > t_{0.05}^{cr}(N_1 - 2) = 2.365$; $F = 41.6 > F_{0.05}^{cr}(f_1 = 1; f_2 = 7) = 5.59$; sum of squares residuals: $\Sigma_1 = 0.0127$; distribution of residues: (Wilk-Shapiro test) $W = 0.786 < W_{0.05}^{cr}(N_1) = 0.829$; straightness feature: $K = 1.02 < K^{thr} = 3.0$ (40)

Statistics of sets Z and H:

$N_1 = 9$, $Z_1^{av} = 2.74 \pm 0.04$; 95% confidence interval (2.64-2.84), $Z_1^{min} = 2.519$, $Z_1^{max} = 2.857$, $S_{Z1} = 0.129$, $\tau^{max} = 0.91 < \tau^{min} = 1.71 < \tau_{0.05}^{cr,2}(N_1) = 2.237 < \tau_{0.05}^{cr,1}(N_1) = 2.392$; Wilk-Shapiro test: $W = 0.803 \approx W_{0.05}^{cr}(N_1) = 0.829$, David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(N_1) = 2.590 < U = [(Z_1^{max} - Z_1^{min})/S_{Z1}] = 2.62 < U2_{0.05}^{cr}(N_1) = 3.552$; (41)

$N_1 = 9$; $H_1^{av} = 1.49 \pm 0.03$; 95% confidence interval: (1.41-1.57), $H_1^{min} = 1.297$, $H_1^{max} = 1.6597$, $S_{H1} = 0.104$, $\tau^{max} = 1.61 < \tau^{min} = 1.88 < \tau_{0.05}^{cr,2}(N_1) = 2.237 < \tau_{0.05}^{cr,1}(N_1) = 2.392$; Wilk-Shapiro test: $W = 0.883 > W_{0.05}^{cr}(N_1) = 0.829$, David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(N_1) = 2.590 < U = [(H_1^{max} - H_1^{min})/S_{H1}] = 3.48 < U2_{0.05}^{cr}(N_1) = 3.552$. (42)

In this case, the linear relationship between Z and H features is characterized by the Chaddock scale as "very close". Now let us check whether this relationship between signs Z and H is preserved (qualitatively and quantitatively) for inactive ($A \leq 35\%$) chemical compounds (Nos. 10-25) from Table 1. The following linear regression was obtained:

$H(Z)_2 = a_{02} + a_{12}\cdot Z$, $N_2 = 16$, $R_2 = 0.54 \pm 0.19$, $R_2^* = 0.57 > R_{0.05}^{cr}(N_2 - 2) = 0.482$; $RMSE(S_2) = 0.059$; sample size sufficient for the validity of the correlation coefficient: $N_{0.05}^{min} = 14$; $a_{02} = 0.68 \pm 0.42$, $a_{12} = 0.25 \pm 0.10$, $t(a_{02}) = 2.65 > t(a_{12}) = 2.51 > t_{0.05}^{cr}(N_2 - 2) = 2.131$; $F = 6.3 > F_{0.05}^{cr}(f_1 = 1; f_2 = 15) = 4.54$; criterion for the significance of the correlation coefficient based on the Fisher normalizing z-transform (taking into account Hotelling's corrections): $u_H = 0.608 > u_{0.05}(N_2) = z_{0.975}\cdot(N_2 - 1)^{-0.5} = 0.506$; sum of squares residuals: $\Sigma_2 = 0.0526$, distribution of residues: (Wilk-Shapiro test): $W = 0.894 > W_{0.05}^{cr}(N_2) = 0.887$; straightness feature: $K = 3.36 > K^{thr} = 3.0$. (43)

Statistics of sets Z_2 and H_2:

$N_2 = 16$, $Z_2^{av} = 2.61 \pm 0.04$; 95% confidence interval: (2.53-2.69), $Z_2^{min} = 2.282$, $Z_2^{max} = 2.833$, $S_{Z2} = 0.153$, $\tau^{max} = 1.44 < \tau^{min} = 2.16 < \tau_{0.05}^{cr,2}(N_2) = 2.523 < \tau_{0.05}^{cr,1}(N_2) = 2.644$; Wilk-Shapiro test: $W = 0.879 \approx W_{0.05}^{cr}(N_2) = 0.887$; David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(N_2) = 3.01 < U = [(Z_2^{max} - Z_2^{min})/S_{Z2}] = 3.60 < U2_{0.05}^{cr}(N_2) = 4.24$; (44)

$N_2 = 16$, $H_2^{av} = 1.33 \pm 0.02$; 95% confidence interval: (1.29-1.37), $H_2^{min} = 1.199$, $H_2^{max} = 1.515$, $S_{Z2} = 0.153$, $\tau^{min} = 1.87 < \tau^{max} = 2.64 < \tau_{0.05}^{cr,2}(N_2) = 2.523 < \tau_{0.05}^{cr,1}(N_2) = 2.644$; Wilk-Shapiro test: $W = 0.904 > W_{0.05}^{cr}(N_2) = 0.887$, David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(N_2) = 3.01 < U = [(H_2^{max} - H_2^{min})/S_{H2}] = 4.45 \approx U2_{0.05}^{cr}(N_2) = 4.24$. (45)

The quality of the regression (43) is noticeably inferior to the quality of the regression (40). Therefore, it is necessary to check if these two regressions can be combined into one linear regression. To do this, we will perform a quantitative statistical comparison of the two regressions. Preliminarily check whether the difference between the variances of residuals (40) and (43) is significant. Let us compare the ratio of the larger residual variance to the smaller residual variance with the tabular value of the F-distribution with $f_1 = N_1 - 2$ and $f_2 = N_2 - 2$ degrees of freedom:

$F = (S_2/S_1)^2 = 1.97 < F_{0.05}^{cr}(f_2 = 14; f_1 = 7) = 3.52$. (46)

The variance of the residuals at the 95% confidence level can be considered homogeneous. In this case, a test [8] can be used to compare the regression coefficients a_{11} and a_{12}:

$$t = |a_{11} - a_{12}|/\{[(N_1 - 2)S_1^2 + (N_2 - 2)S_2^2)/(N_1 + N_2 - 4)]\cdot[1/(N_1 - 2)/S_{Z1}^2 + 1/(N_2 - 2)/S_{Z2}^2]\}^{0.5} =$$

$$= 2.87 > t_{0.05}^{cr}(N_1 + N_2 - 4) = 2.080. \quad (47)$$

Here $S_{Z1} = 0.129$ (41) and $S_{Z2} = 0.153$ (44) are the standard deviations of the explanatory variables in regressions (40) and (43). It follows from inequality (47) that the estimates of the regression coefficients (40) and (43) differ significantly. Thus, it can be agreed that the Z and H relationships are

significantly different for bioactive drugs and weakly active (or inactive) chemical compounds (at the 95% confidence level). Therefore, it can be assumed that there is a structural shift in the $H(Z)$ relationships during the transition from chemical compounds with pronounced bioactivity to many chemical compounds that do not have significant radioprotective activity (Fig. 5). Such a structural shift can be considered as an indicator that is associated with the biological action of chemical compounds in Table 1. This indicator makes it possible to presumably separate bioactive chemical compounds from inactive or weakly active preparations.

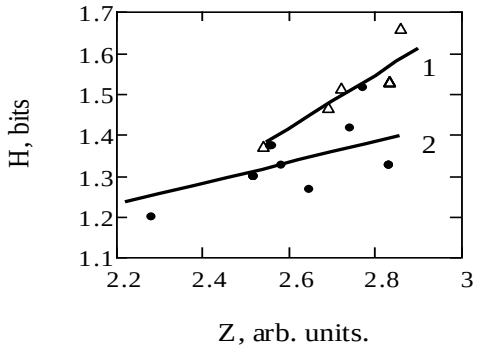

Fig. 5. Correlation field and scatter diagrams. Δ - bioactive chemical compounds. 1 - regression line (40). • - inactive or weakly active chemical compounds. 2 - regression line (43).

Let us also compare the correlation coefficients of regressions (40) and (43). For this purpose, we use statistics [8,17], which has an approximately normal distribution:

$$\Lambda = |z_1 - z_2|/[(N_1 - 1)^{-1} + (N_2 - 1)^{-1}]^{0.5}. \quad (48)$$

Here we used the normalizing Fisher z-transform $z = 0.5\cdot\ln[(1 + |R^*|)/(1 - |R^*|)]$ for the correlation coefficients R_1^* (40) and R_2^* (43): $z_1 = 1.74$, $z_2 = 0.65$. From relation (48) we obtain the following inequality:

$$\Lambda = 2.49 > \Lambda_{0.05}^{cr} = 1.96. \quad (49)$$

$\Lambda_{0.05}^{cr}$ is a critical quantity at significance of $\alpha = 0.05$ [8]. At the 95% confidence level, inequality (49) establishes a significant difference between the correlation coefficients. That is, the null hypothesis of equality of correlation coefficients is rejected. It is also possible to estimate the summary value of the correlation coefficient:

$$z^{sum} = [z_1\cdot(N_1 - 3) + z_2\cdot(N_2 - 3)]/(N_1 + N_2 - 6) = 0.994. \quad (50)$$

The result (50) gives an estimate of the significance of the difference from zero of the summary correlation coefficient z^{sum} for the two samples:

$$\Lambda = z^{sum}\cdot(N_1 + N_2 - 6)^{0.5} = 4.33 > \Lambda_{0.05}^{cr} = 1.96. \quad (51)$$

Inequality (51) indicates that there is a significant relationship between the electronic Z and informational H features for samples N_1 and N_2. Second, according to test (47), these relationships have a statistically significant different intensity of change (Fig. 5). Application of the Chow test (13) does not deny the presence of a structural break in the relationship between H and Z characteristics. Indeed, using the sums of squares of the remainder (40) and (43), as well as relation (13), we obtain the following inequality:

$$F = 13.73 > F_{0.05}^{cr}(f_1 = m + 1; f_2 = N - 2m - 2) = 3.47. \quad (52)$$

Here $\Sigma = 0.1507$ is the sum of the squares residuals for total regression (for the sample $N = N_1 + N_2$). Inequality (52) assumes that at the significance level $\alpha = 0.05$ and the number of degrees of freedom $f_1 = 2$ and $f_2 = 21$, the regressions (40) and (43) differ significantly and cannot be combined into one linear regression. Thus, an additional statistical indicator (structural shift) characterizes the difference between bioactive drugs and weakly active chemical compounds.

Now let us perform a statistical analysis of the relationship between changes in the molecular structure of chemical compounds and the variability of radioprotective activity (radiation dose of 700 R) for tryptamine derivatives with substitution in the carbon side chain (Table 4). The analysis showed that a significant factor influencing the bioactivity of molecules is the geometric size of the substituents in the side chain. The five-dimensional (L, B_1 - B_4) A.Verloop steric parameters were used for the substitutes [34]. Only parameters B_4 and L were statistically significant. However, for the sample (Table 3), these parameters of substituents are closely linearly interrelated. That is, they are collinear. Therefore, one of the parameters must be excluded from the regression. In what follows, only one steric parameter B_4 will be used. We assume that the explanatory molecular features are included the regression equation additively.

For chemical compounds Nos. 26–36 (tryptamine was added to the sample), the relationship of bioactivity (Act_{exp},%) with each of the explanatory

features Q_5 and B_4 was analyzed. The analysis showed that a statistically significant relationship is either missing or the quality of regression is relatively low. For example, when using the complex molecular trait Q_5 in regression, we obtain the following value of the correlation coefficient: $|R| = 0.22 < R_{0.05}^{cr}(N - 2) = 0.602$, $RMSE = 0.11$, $N = 11$.

Table 3
Substitution in the side chain. The experimental data on the radioprotective effect of drugs were taken from [2].

No	Chemical compounds	ε_{unoc}, eV	q_5	Q_5, arb. units.	Z, arb. units	H, *bits*	B_4, arb. units	MR, cm^3/mol	Dose, mg/kg	Act_{exp} %	$Act9$, %
26	N-Acetyl-5-methoxytrypt-amine $R_3 = (CH_2)_2N(H)COCH_3$ $R_5 = OCH_3$	-0,820	-0.2099	-0.3757	2.73	.53	5.61	7.87	50-75	0	3.9
27	N,N'- Dimethyltryptamine $R_3 = (CH_2)_2N(CH_3)_2$	-0.595	-0.0080	-0.0532	2.47	1.27	4.80	1.03	55	6.6	6.9
28	N- Ureidotryptamine $R_3 = (CH_2)_2NHC=O(NH_2)$	-0.544	-0.0010	-0.0172	2.70	1.49	5.61	1.03	68	0	0
29	N- Thioureido-5-methoxytryptamine $R_3 = (CH_2)_2NHC=S(NH_2)$ $R_5 = OCH_3$	-1,110	-0.2009	-0.4975	2.87	1.69	6.18	7.87	75	0	0
30	N- Ureido-5-methoxytryptamine $R_3 =(CH_2)_2NHC=O(NH_2)$ $R_5 = OCH_3$	-1.130	-0.2056	-0.5124	2.79	1.64	5.61	7.87	72	0	8.1
31	α- Methyltryptamine $R_3 = CH_2CH(CH_3)NH_2$	-0.610	-0.0075	-0.0528	2.52	1.30	4.92	1.03	55	0	4.9
32	α- Methyl-5-chloro-tryptamine $R_3 = CH_2CH(CH_3)NH_2$, $R_5 = Cl$	-1,025	-0.2165	-0.4766	2.74	1.49	4.92	6.03	65	12.5	18.1
33	α- Methyl-5-fluoro-tryptamine $R_3 = CH_2CH(CH_3)NH_2$,$R_5 = F$	-0.916	-0.2211	-0.5191	2.74	1.49	4.92	0.92	60	30.0	19.4
34	β- Methyltryptamine $R_3 = CH(CH_3)CH_2NH_2$	-0.707	-0.0079	-0.0628	2.52	1.30	4.92	1.03	55	0	5.2
35	N- Monomethyltryptamine $R_3 =(CH_2)_2NHCH_3$	-0.601	-0.0079	-0.0539	2.52	1.30	5.08	1.03	50	0	2.4
36	Tryptamine	-0.614	-0.0079	-0.0548	2.58	1.33	3.84	1.03	50	23.0	22.3

At the same time, the explanatory variable B_4 has a "noticeable" (according to the Chaddock scale, the correlation coefficient falls within the range of $0.5 \leq |R| \leq 0.7$ [35,36]:

$$Act8/100 = b_0 + b_1 \cdot B_4,$$

$N = 11$, $R_8 = - 0.62 \pm 0.21$, $|R_8^*| = 0.64 > R_{0.05}^{cr}(N - 2) = 0.602$, sample size sufficient for the validity of the correlation coefficient: $N_{0.05}^{min} = 10$; $RMSE_8 = 0.089$; criterion for the significance of the correlation coefficient based on the Fisher normalizing z-transform (taking into account Hotelling's corrections): $u_H = 0.666 > u_{0.05}(N) = z_{0.975} \cdot (N - 1)^{-0.5} = 0.619$; $b_0 = 0.62 \pm 0.24$, $b_1 = - 0.11 \pm 0.05$, $t(b_0) = 2.616 > |t(b_1)| = 2.356 > t_{0.05}^{cr}(N - 2) = 2.262$; $F = 5.55 > F_{0.05}^{cr}(f_1 = 1; f_2 = 9) = 5.12$; the residuals are normally distributed: $W = 0.851 > W_{0.05}^{cr}(N) = 0.850$; $\Sigma_8 = 0.0715$, $AIC_8 = -4.8541$, $SC_8 = -4.6218$, $SS_8 = 0.0267$. (53)

Checking the relationship of the regression residuals (53) with the Q_5 indicator showed the presence of a significant linear relationship: $|r| = 0.65 > R_{0.05}^{cr}(N - 2) = 0.602$; $F = 6.72 > F_{0.05}^{cr}(f_1 = 1; f_2 = 9) = 5.12$.

Therefore, it is possible to construct a two-factor regression that takes into account the combined influence of explanatory variables B_4 and Q_5 on the variability of bioactivity. Two-factor regression was found to be significant:

$$Act9/100 = b_0 + b_1 Q_5 + b_2 \cdot B_4,$$

$N = 11$, $m_1 = 2$, $R_9 = 0.85 > R_{0.95}^{cr}(\nu = 8; m = 2) = 0.726$ [26], $R_9^2 = 0.72$, $R_9^{*2} = 0.69$, $RMSE_9 = 0.064$, $b_0 = 0.82 \pm 0.18$, $b_1 = -0.31 \pm 0.10$, $b_2 = -0.16 \pm 0.04$, $t(b_0) = 4.49 > |t(b_2)| = 4.33 > |t(b_1)| = 3.08 > t_{0.05}^{cr}(f = 8) = 2.306$; $F = 10.13 > F_{0.05}^{cr}(f_1 = 2; f_2 = 8) = 4.46$; the residuals are normally distributed: $W = 0.914 > W_{0.05}^{cr}(N) = 0.850$; $\Sigma_9 = 0.0327$, $AIC_9 = -5.4546$, $SC_9 = -5.1643$, $SS_9 = 0.0201$. (54)

Statistics of sets Q_5 and B_4:

Q_5: $N = 11$, $Q_5^{av} = -0.24 \pm 0.07$; 95% confidence interval: (-0.40, -0.09); $Q_5^{min} = -0.5191$, $Q_5^{max} = -0.0172$, $S_Q = 0.226$, $\tau^{max} = 0.98 < \tau^{min} = 1.23 < \tau_{0.05}^{cr,2}(N) = 2.343 < \tau_{0.05}^{cr,1}(N) = 2.484$; David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(N) = 2.74 > U = [(Q_5^{max} - Q_5^{min})/S_Q] = 2.22 < U2_{0.05}^{cr}(N) = 3.80$;

B_4: $N = 11$, $B_4^{av} = 5.13 \pm 0.18$; 95% confidence interval: 4.72 - 5.54; $B_4^{min} = 3.84$, $B_4^{max} = 6.18$, $S_B = 0.612$, $\tau^{max} = 1.72 < \tau^{min} = 2.11 < \tau_{0.05}^{cr,2}(N) = 2.343 < \tau_{0.05}^{cr,1}(N) = 2.484$; David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(N) = 2.74 < U = [(B_4^{max} - B_4^{min})/S_B] = 3.83 \approx U2_{0.05}^{cr}(N) = 3.80$. (55)

It is possible to establish the statistical importance of the additional attribute Q_5 in the regression equation for the purpose of explaining the variability of bioactivity. For this we use the following relation [8]:

$$F = (R_9^{*2} - R_8^{*2})(N - m - 1)/(m - m_1)/(1 - R_9^{*2}) = 6.45 > F_{0.05}^{cr}(f_1 = m - m_1; f_2 = N - m - 1) = 5.12. \quad (56)$$

Since inequality (56) holds, the explanatory variables together have a significant effect on bioactivity. This result does not contradict the quality criteria (53) and (54).

However, the presence of collinearity between the explanatory variables should be checked. Since the regression residuals (54) are normally distributed (Wilk-Shapiro test: $W = 0.860 > W_{0.05}^{cr}(N) = 0.850$), the Farrar-Glauber relation can be used to quantify the collinearity of the variables (32):

$$\chi^2 = 2.021 < t_{0.05}^{cr}(f = 1) = 3.841. \quad (57)$$

Additionally, we also use the t - criterion (12):

$$t = 1.55 < \chi_{0.05}^{2,cr}(f = N - m) = 2.262. \quad (58)$$

The correlation coefficient between the explanatory variables is $r_{1,2} = -0.46$; $m = 2$ is the number of variables. Inequalities (57) and (58) indicate that collinearity between the explanatory variables is statistically insignificant.

The standardized regression coefficients (54), in accordance with (35), turned out to be as follows:

$$b_1^* = S_Q b_1/S_{Act} = -0.662, \quad b_2^* = S_B b_2/S_{Act} = -0.918. \quad (59)$$

Here $S_{Act} = 0.1075$ is the standard deviation of the observed bioactivity values $Act/100$. Using the ratio (52) one can get an approximate estimate of explanatory variables contributions in the bioactivity variability:

$$R_{app}^2 = b_1^* \cdot r_{Q,Act} + b_2^* \cdot r_{B,Act} = 0.150 + 0.560 = 0.71. \quad (60)$$

Here $r_{Q,Act} = -0.230$ and $r_{B,Act} = -0.617$ are the correlation coefficients of the explanatory factors Q_5 and B_4 with the observed bioactivity. The estimate of the approximate coefficient of determination $R_{appr}^2 = 0.71$ (60) is very close to the value $R_9^2 = 0.72$ (54). The proportional contributions to the bioactivity variation for the explanatory variables Q_5 and B_4 will be the next 15% and 56%, respectively. Thus, the influence of steric factor B_4 on the variation in bioactivity of preparations is greater than the influence of electronic factor Q_5.

Let us again test the hypothesis about the validity of the classification rule. Let us apply equation (54) to the N,N'-dimethyl-5-methoxytryptamine molecule (dose: 29 mg/kg; survival rate: 0%). This drug was not included in the initial sample. For this chemical compound, the value of the complex electron factor was calculated by the quantum mechanical method: $Q_5 = -0.2925$ arb. units. Using equation (2) in Figure 2A, we obtain an independent estimate of the factor $Q_{52}(A) = -0.28$ arb. units (for the substituent CH_3O, the electron affinity was determined by the ion impact method [12]: 1.5 ± 0.5 eV). For the CH_3O substituent, the factor $Z = 2.6$ arb. units, which corresponds to the value $Q_{52}(Z) = -0.28$ arb. units (line 2, Fig.2B). Steric factor of the substituent $B_4 = 4.8$ arb. units [35]. The values of Q_5 and B_4 are in the range of acceptable values (55). From equation (54) it follows that the N, N'-dimethyl-5-methoxytryptamine molecule is supposed to have a weak radioprotective effect: $Act9 = 14.3\%$ (12.0%). The survival rate obtained using the estimate of the variable $Q_{52} = -0.28$ arb. units is indicated in

parentheses. Table 3 gives the bioactivity values calculated using equation (54).

The regression equation (54) explains only 72% of the variance of the bioactivity for chemical compounds (Table 4**)**. The rest changes, apparently, are associated with other hidden or unaccounted for factors and random phenomenon. Therefore, a model was tested that takes into account the combined action of three factors Q_5, MR^2 and B_4:

$$Act10/100 = b_0 + b_1 \cdot Q_5 + b_2 \cdot B_4 + b_3 \cdot MR^2,$$

$N = 11$, $R_{10} = 0.92 > R_{0.95}^{cr}(\nu = N - m - 1; m = 3) = 0.807$ [26], $R_{10}^2 = 0.847$, $R_{10}^{*2} = 0.79$, $RMSE_{10} = 0.049$, $b_0 = 0.59 \pm 0.17$, $b_1 = -0.49 \pm 0.11$, $b_2 = -0.11 \pm 0.03$, $b_3 = -0.0025 \pm 0.001$, $|t(b_1)| = 4.62 > t(b_0) = 3.47 > |t(b_2)| = 3.33 > |t(b_3)| = 2.50 > t_{0.05}^{cr}(f = 7) = 2.365$; $F = 13.28 > F_{0.05}^{cr}(f_1 = m; f_2 = N - m - 1) = 4.35$; the regression residuals are normally distributed (Wilk-Shapiro test: $W = 0.884 > W_{0.05}^{cr}(N) = 0.850$); $\Sigma_{10} = 0.0172$, $AIC_{10} = -5.9153$, $SC_{10} = -6.0971$, $SS_{10} = 0.0164$. (61)

Here $m = 3$ is the number of explanatory variables. Comparison of quality tests of models (54) and (61) demonstrates the deterioration of all criteria. That is, it indicates the qualitative advantage of the model (61).

Statistics of set MR^2:

$N = 11$, $MR^{2,av} = 20.94 \pm 8.60$; 95% confidence interval: (1.78 - 40.10); $MR^{2,min} = 0.846$, $MR^{2,max} = 62.25$; $S_{MR2} = 28.52$; $\tau^{min} = 0.71 < \tau^{max} = 1.45 < \tau_{0.05}^{cr,2}(N) = 2.343 < \tau_{0.05}^{cr,1}(N) = 2.484$; David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(N) = 2.740 > U = [(MR^{max} - MR^{min})/S_{MR}] = 2.15 < U2_{0.05}^{cr}(N) = 3.800$; (62)

statistics of set $Act \equiv Act10/100\%$:

$N = 11$, $Act^{av} = 0.066 \pm 0.030$; 95% confidence interval: (-0.007, 0.138); $Act^{min} = 0$, $Act^{max} = 0.30$, $S_{Act} = 0.108$, $\tau^{min} = 0.61 < \tau^{max} = 2.17 < \tau_{0.05}^{cr,2}(N) = 2.343 < \tau_{0.05}^{cr,1}(N) = 2.484$; David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(N) = 2.740 < U = [(Act^{max} - Act^{min})/S_{Act}] = 2.78 < U1_{0.05}^{cr}(N) = 3.800$. (63)

Using MR instead of MR^2 as an explanatory variable does not practically change the results (61). The following standardized regression coefficients were obtained (61):

$b_1^* = S_Q b_1/S_{Act} = -1.02$, $b_2^* = S_B b_2/S_{Act} = -0.65$,

$b_3^* = S_{MR2} b_3/S_{Act} = -0.66$. (64)

The contributions sharing of explanatory variables is estimated from a ratio similar to (30):

$$R_{appr}^2 = b_1^* \cdot r_{Q,Act} + b_2^* \cdot r_{B,Act} + b_3^* \cdot r_{MR2,Act} = 0.235 + 0.402 + 0.217 = 0.853. \quad (65)$$

Correlation coefficients: $r_{Q,Act} = -0.230$, $r_{B,Act} = -0.617$ and $r_{MR2,Act} = -0.33$. The value of the approximate coefficient of determination $R_{appr}^2 = 0.853$ is very close to the value $R_{10}^2 = 0.847$ (61). From relation (65) it follows that the main contribution to the regression is made by the variable B_4. The high statistical significance of the geometric size B_4 of the side chain substituent indicates that close proximity of the drug molecule to the target substrate is a necessary condition for the manifestation of the biological activity of an exogenous molecule. The use of the explanatory variable MR in the regression (61) instead of the explanatory variable MR^2 did not lead to a significant improvement in the quality of the regression.

The analysis of bioactivity (Table 4) demonstrates that the complementarity of molecules to the receptor is a necessary condition for the manifestation of biological activity by a chemical compound. The influence of the size B_4 of substituents in the side chain can reduce the role of the feature Q_5 in predicting the radioprotective efficacy of a chemical compound. For example, chemical compounds Nos. 29, 30 and 32 from Table 4 have a rather large negative value of the Q_5 parameter. Moreover, the factor-sign Q_5 is significantly greater (in absolute value) than the threshold value Q^{thr} (Fig. 3B). However, the steric factor B_4 for substituents in the side chain is noticeably higher than for the hydrogen atom 3.84 arb. units. A decrease in the size B_4 of the substituent (for example, the molecule No. 27) is accompanied by the appearance of a weak radioprotective effect of the drug, with a relatively small (in absolute value) value of the factor $Q_5 = -0.0532$ arb. units. This value is close to the value obtained for tryptamine (Table 1). At the same time, this value differs markedly from the expected threshold value Q^{thr}. Apparently, the substituted tryptamine molecules carry out complementary binding to the receptor, including through the formation of a charge transfer complex. It is well known that charge transfer complexes are formed by short-range interaction forces, and the position of the MO of the ε_{unoc} level on the energy scale plays an important role. Therefore, the bioactivity of molecules is very "sensitive" to the size of the substituent in the side chain and its compatibility

with the biological object. For example, the values of the parameter Q_5 for molecules No. 30 and No. 33 are very close and equal to -0.5124 and -0.5191 arb. units, respectively. However, due to the larger size of the substituent in the side chain of molecule No. 30, the radioprotective effect completely disappears. A similar situation also occurs for the N-thioureido-5-methoxytryptamine molecule (No. 29). The molecule has a negative value of the factor Q_5 = - 0.4975 arb. units, far from the threshold value of Q^{thr} and in absolute value greater than that of mexamine (No. = 6 in Table 1; Q_5 = - 0.2767 arb. units). However, the largest value of the steric parameter is B_4 = 6.18 arb. units for the N-thioureido-5-methoxytryptamine molecule leads to a decrease in bioactivity in comparison with the mexamine molecule (Table 1). It can be assumed that intermolecular interactions determine not only the interaction of the molecule with the active center of the biophase, but also the direction of the translational movement of the molecule to the receptor. At the same time, short-range interactions are involved in the quasi-chemical binding of the molecule to the receptor. Moreover, the limiting factor in the manifestation of bioactivity is the complementarity of the geometric size B_4 of the substituent in the side chain. The importance of the steric size B_4 of the substituent for the radioprotective effect is also indicated by the comparison of preparations No. 1 and 3 from Table 1 with preparations No. 32 and 33 from Table 4. The values of the Q_5 calculated by the quantum mechanical method for these compounds are very close, but the geometric sizes of the B_4 is significantly different. Indeed, for a hydrogen atom B_4 = 3.84 arb. units (Nos. 1 and 3), and for CH_2 group (Nos. 32 and 33) 4.92 arb. units, respectively. As a result, the compatibility of the drug when interacting with the target appears to be hampered. Therefore, the radioprotective effect of drugs is significantly reduced. Therefore, it can be assumed that the optimal size B_4 should be in the range of values less than 4.0 arb. units. The optimality of the size of the aminoethyl structure of the side chain of tryptamine was also indicated in article [2]. For the sample (Table 4), an estimate was made of the relationship between molecular features Z and H:

$H(Z)_3 = a_{03} + a_{13}\cdot Z$, $N_3 = 11$, $R_3 = 0.98 \pm 0.01$, $|R_3^*| = 0.982 > R_{0.05}^{cr}(N_3 - 2) = 0.602$; $RMSE(S_3) = 0.031$; criterion for the significance of the correlation coefficient based on the Fisher normalizing z-transform (taking into account Hotelling's corrections): $u_H = 2.168 > u_{0.05}(N_3) = z_{0.975}\cdot(N_3 - 1)^{-0.5} = 0.620$; size sufficient for the validity of the correlation coefficient: $N_{0.05}^{min} < 5$; $a_{03} = -1.41 \pm 0.19$, $a_{13} = 1.08 \pm 0.07$, $t(a_{13}) = 14.8 > |t(a_{03})| = 7.31 > t_{0.05}^{cr}(f = N_3 - 2) = 2.262$; $F = 218.2 > F_{0.05}^{cr}(f_1 = 1; f_2 = 9) = 5.12$; straightforwardness feature: $K = 0.66 < K^{thr} = 3.0$. (66)

Statistics of sets Z and H:

Z: $Z^{av} = 2.65 \pm 0.04$; the 95% confidence interval is (2.56 - 2.74); $Z^{min} = 2.47$, $Z^{max} = 2.87$, $S_{Z3} = 0.135$, $\tau^{min} = 1.33 < \tau^{max} = 1.63 < \tau_{0.05}^{cr,2}(N_3) = 2.343 < \tau_{0.05}^{cr,1}(N_3) = 2.484$; Wilk-Shapiro normality test: $W = 0.907 > W_{0.05}^{cr}(N_3) = 0.850$; David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(N_3) = 2.740 < U = [(Z^{max} - Z^{min})/S_{Z3}] = 2.96 < U2_{0.05}^{cr}(N_3) = 3.800$;

H: $H^{av} = 1.44 \pm 0.05$; 95% confidence interval is: (1.34 - 1.54); $H^{min} = 1.27$, $H^{max} = 1.69$, $S_{H3} = 0.147$, $\tau^{min} = 1.15 < \tau^{max} = 1.70 < \tau_{0.05}^{cr,2}(N_3) = 2.343 < \tau_{0.05}^{cr,1}(N_3) = 2.484$; Wilk-Shapiro normality test: $W = 0.882 > W_{0.05}^{cr}(N_3) = 0.850$, David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(N_3) = 2.740 < U = [(H^{max} - H^{min})/S_{H3}] = 2.86 < U2_{0.05}^{cr}(N_3) = 3.800$. (67)

Since the variances S_1^2 (40) and S_3^2 (61) do not differ significantly:

$$F = (0.043/0.031)^2 = 1.92 <$$

$$F_{0.05}^{cr}(f_1 = N_1 - 2; f_2 = N_2 - 1) = 3.37, \quad (68)$$

then we can use relation (47) to estimate the statistical significance of the difference between regression coefficients a_{11} (40) and a_{13} (63):

$$t = |a_{11} - a_{13}|/\{[((N_1 - 2)S_1^2 + (N_3 - 2)S_3^2)/(N_1 + N_3 - 4)]\cdot[1/(N_1 - 2)/S_{Z1}^2 + 1/(N_3 - 2)/S_{Z3}^2]\}^{0.5} =$$

$$= 2.50 > t_{0.05}^{cr}(f = N_1 + N_3 - 4) = 2.12. \quad (69)$$

Thus, it follows from inequality (69) that the regression coefficients a_{11} and a_{13} are different for bioactive drugs and weakly active or inactive chemical compounds (Table 3). That is, a structural shift in the relationship of molecular factors $H(Z)$ takes place in this case too.

4 Conclusion

Modeling of the biological activity of substituted tryptamines showed that in the series of indolylalkylamines there is a statistically significant relationship between the structure of the initial molecules (electronic and dimensional properties of substituents) and their radioprotective effect.

As follows from equation (61), the more negative the feature Q_5, and the smaller the geometric size of

the substituents in the side chain, as well as the bulk size of the substituent at position R_5, the higher the antiradiation activity of the tryptamine-based drug. Apparently, the interaction of the molecule due to electrostatic forces and donor-acceptor transfer processes should be sufficiently strong (the important role of threshold), and the active region of the molecule should be complementary to the receptor for efficient electron transfer.

References:

[1] Bacq Z.M., Alexander P., *Fundamentals of Radiobiology*, Pergamon Press, 1961.
[2] Zherebchenko P.G., *Antiradiation Properties of Indolylalkylamines*, Moscow, Atomizdat, 1971 (in Russian).
[3] Renson J., *Arch. Intern. Physiol. Biochem*., Vol.68, 1960, p.531.
[4] Dukor P., Schuppli R., *Experientia*, Vol.17, 1961, p.257.
[5] Dukor P., *Strahlentherapie*, Bd.117, 1962, p.330.
[6] Supek Z., Dandči M., Lovašen Z., *Inter. J. Rad. Biol.*, Vol.4, No.1, 1961, p.111.
[7] Schembelev G.A., Ustynyuk V.M., Mamaev V.M. and other, *Quantum-Chemical Methods for Calculating Molecules*, Moscow, MSU, 1980 (in Russian).
[8] Förster E., Rönz B., *Metohden der Korrelations- und Regressionsanalyse*, Verlag Die Wirtschaft Berlin, 1979.
[9] Mukhomorov V.K., *Radiobiology*, Vol.26, No 4, 1986 pp. 557-559 (in Russian).
[10] Mukhomorov V.K., *Chem. Rapid Commun.*, Vol.1, No1, 2013, pp.15-19.
[11] Szent–Györgyi Albert, *Introduction to a Submolecular Biology*, Acad. Press Inc., New York, 1960.
[12] Gurevich L.V., Karachevtsev G.V., Kondrat'ev V.N., Lebedev Yu.A., Medvedev V.A., Potapov V.K., Khodeev Yu.S., *The Energies of Breaking Chemical Bonds. Ionization Potentials and Electron Affinity*, Moscow, 1974 (in Russian).
[13] Smirnov B.M., *Atomic Collisions and Elementary Processes in Plasma*, Moscow, 1968 (in Russian).
[14] Wentworth W.E., Chen E., Steelhammer J.C., *J. Phys. Chem*., Vol.72, 1967, p.2671. 1967.
[15] Zaitsev G.N., *Mathematics in Experimental Botany*,. Moscow, 1990 (in Russian).
[16] Zachs L, *Statistische Auswertungsmetoden*, Berlin. 1972.
[17] Johnson N.L., Leon F.C., *Statistical and Experimental Design. In Engineering and the Physical Sciences*, Vol. 1, JohnWiley & Sons, New York, 1977.
[18] Kobzar A.I., *Applied Mathematical Statistics. For engineers and scientists*, Moscow, 2016 (in Russian).
[19] Chaddock R.E., *Principles and Methods of Statistics*, Boston, New York, 1925.
[20] Draper N.R., Smith H., *Applied Regression Analysis*, Vol.1, John Wiley & Sons, New York, 1981.
[21] Chow Gregory C., Tests of Equality Between Sets of Coefficients in Two Linear Regressions, *Econometrica*, Vol. 28, 1960, pp. 591-605.
[22] Szent–Györgyi Albert, *Bioenergetics*, Acad. Press Inc., New York,1967.
[23] Linnik Yu.V., *The Method of Least Squares and the Foundations of the Mathematical and Statistical Theory of Observation Processing*, Moscow, 1962 (in Russian).
[24] Sturges H.A., *J. Amer. Statist. Assoc*., Vol. 21, 1926, pp.65-66.
[25] *Handbook of Applicable Mathematics*, Vol.VI, Part A, Chief Editor Walter Ledermann, John Willey&Sons, Chichester-New York-Brisbane, 1984.
[26] Likeš J., Laga J., *Zakladni Statistike Tabulky*, Praha, 1978.
[27] Akaike Hirotogu, A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, Vol.19, No.6, 1974, pp. 716-723.
[28] Schwarz G., Estimating the dimension of model, *Annals of Statistics*, Vol.6, 1978, pp.461-464.
[29] Hansch C., Leo A., *Substituent Constants for Correlation Analysis in Chemistry and Biology*, John Wiley & Sons, New York, Chichester, Bribane, Toronto, 1979.
[30] Hansh C., Leo A., Unger S. H., Ki Hwan Kim, Nikaitani D., Lien E.J., *J. Med. Chem.*, Vol.16, No.11, 1073, p.1207.
[31] Leo A., Hansch C., Elkins D., *Chem. Rev*., Vol.71, 1971, pp.525-616.
[32] Mukhomorov V.K.**,** The New model of Carcinogenic Activity, *WSEAS Transaction on Biology and Medicine*, Vol. 18, 2021, pp.150-169.
[33] Verloop A., Hoogenstraaten W., Tipker J., *Development and Application of New Steric Substituent Parameters in Drug Design. In: Drug Design*, Ed. F.J. Ariens, Acad. Press, New-York, 1976, Vol.7, pp.165-207.
[34] Baraz V.R. *Correlation-Regression Analysis of the Relationship between Commercial Performance Indicators Using EXCEL Programs*, Yekaterinburg, 2005, (in Russian).
[35] https://math.semestr.ru/corel/cheddok.php

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0
https://creativecommons.org/licenses/by/4.0/deed.en_US