

The New Model of Carcinogenic Activity of Polycyclic Hydrocarbons

VLADIMIR K. MUKHOMOROV

Physical Department
University "Federico II"
Via Cintia, I-80126, Napoli
ITALY

Abstract: - A model is proposed that allows one to interpret the carcinogenic properties of polycyclic chemical compounds. Electronic, informational and structural molecular factors that characterize the molecule as a whole are proposed as explanatory variables. The factors limiting the carcinogenic activity of polycyclic compounds are analyzed. The model fully interprets all observable data that were used to support previous early models.

Key-Words: - Polycyclic, carcinogenicity, threshold, information function, electronic factor, statistical criteria, intermolecular interactions, electron affinity, structural factor.

Received: April 22, 2021. Revised: July 10, 2021. Accepted: August 15, 2021. Published: September 29, 2021.

1 Introduction

A special class of carcinogenic chemical compounds are polycyclic aromatic hydrocarbons, which differ from many other known carcinogens. Polycyclic aromatic hydrocarbons are powerful carcinogenic compounds found in the environment, found in various petroleum products and combustion products. In addition, polycyclic aromatic hydrocarbons are usually formed by pyrolysis of organic substances, as a result of smoking tobacco, certain cooking procedures. Human exposure to polycyclic aromatic hydrocarbons is extensive. Polycyclic aromatic hydrocarbons have several characteristics that distinguish them from many of the more well-known carcinogens. The effective dose is in the order of micrograms, and they have been found to induce tumors in almost all types of tissues and animals in which they have been tested [1,2].

2 Problem Formulation

Attempts have been made repeatedly to study the relationship of carcinogenic activity with the electronic properties of polycyclic aromatic hydrocarbons. For example, in paper [3], the energy of electron localization in hydrocarbons was calculated. Using the parameters of Muller [4], an attempt was made to link the carcinogenic effect with the variation of hyperconjugation in aromatic molecules. The most successful modeling of the relationship between carcinogenicity and the molecular structure of hydrocarbons was performed

by A. Pullman and B. Pullman [5]. These authors suggested linking carcinogenic activity to the electronic properties of two regions of polycyclic molecules. They called these two regions *K* and *L* areas. This model was developed in the works of other researchers, for example, in [6,7]. The authors tried to establish the relationship between the carcinogenic effect of polycyclic hydrocarbons and the reactivity of the *K* and *L* regions of the molecules. Within the framework of this model, the variation in the carcinogenic activity of hydrocarbons is due to changes in the reactivity of the *K*-region. However, the correlation with the carcinogenic activity of a wide range of polycyclic hydrocarbons was not entirely satisfactory. The situation could not be corrected by more detailed quantum-chemical calculations [1]. The difficulties of the model [5] required the introduction of additional reaction regions into the model the *M* and *B* regions of the molecule. However, the relationship of electronic parameters with the carcinogenic activity of polycyclic aromatic hydrocarbons is not as satisfactory as expected.

Even if we additionally ascribe a favorable electron density to anthracenoid meso carbon atoms (*L*-region) and positions in the angular benzene ring, the most sensitive to metabolic oxidation (*M*-region) [1,2]. Even in the early works of Ch. Coulson [8], having considered the uncertainties of the foundations of the model, on which the calculation of numerical values depends, he made a pessimistic conclusion about the relationship between the electronic properties of polycyclic hydrocarbons and

their carcinogenicity. The computational procedure of this model is complex and laborious for rapid analysis or prediction of the possible carcinogenicity of a substance. This model assumes the use of special computer programs for performing quantum chemical calculations in order to obtain the electronic indices of molecules and their subsequent analysis. In addition, the available models often leave aside such questions as the role of exogenous molecules in the mechanism of carcinogenesis of polycyclic aromatic hydrocarbons, or the problem of limiting the carcinogenic activity of substances.

3 Problem Solution

Table 1
Additive components of the pair energy of interaction of tetramethyl uric acid with polycyclic hydrocarbons [13,14].

№	Chemical compound	Gross-formula	Contributions to the energy of pair interaction, kcal/mol				Z, arb. units	H, bits
			E_{elect}	E_{pol}	E_{disp}	E_{rep}		
1	Dibenz-1,2,3,4-pyrene	C ₂₄ H ₁₄	-1.93	-1.11	-10.37	4.95	2.90	0.95
2	Dibenz-1,2,5,6-anthracene	C ₂₂ H ₁₄	-1.09	-0.79	-8.23	3.88	2.94	0.94
3	Anthantrene	C ₂₂ H ₁₂	-1.34	-0.93	-8.72	4.79	2.94	0.94
4	Perylene	C ₂₀ H ₁₂	-1.35	-0.86	-7.80	4.25	2.88	0.96
5	Benz-1,2-pyrene	C ₂₀ H ₁₂	-1.77	-0.96	-9.23	4.51	2.88	0.96
6	Benz-3,4-pyrene	C ₂₀ H ₁₂	-1.38	-0.86	-8.64	4.48	2.88	0.96
7	Pyrene	C ₁₆ H ₁₀	-1.46	-0.78	-7.76	4.12	2.85	0.96
8	Chrysene	C ₁₈ H ₁₂	-1.18	-0.80	-7.92	4.15	2.80	0.97
9	Benz-1,2-anthracene	C ₁₈ H ₁₂	-1.18	-0.80	-7.92	4.15	2.80	0.97
10	Фенантрен	C ₁₄ H ₁₀	-1.57	-0.65	-7.13	3.77	2.75	0.98
11	Anthracene	C ₁₄ H ₁₀	-1.27	-0.79	-6.84	3.67	2.75	0.98

The model establishes the relationship of the carcinogenic activity of classical polycyclic aromatic compounds with the molecular factors Z (electronic factor), H (information factor) and γ (structural factor) which were used in the analysis of the carcinogenic activity of metabolites [10] and in predicting the biological activity of chemical compounds [11].

An alternative model proposed here uses molecular factors, which, in contrast to the Pullman model, characterize not individual regions of molecules, but the molecule as a whole. Some additional ideas about the physical meaning of the molecular factor attribute Z can be obtained by analyzing the additive components of the total energy of the pair intermolecular interaction of chemical compounds with model molecular systems. It is known, for example, that polycyclic aromatic hydrocarbons are capable of being

Within the framework of the model [5], the conclusion that osmium tetroxide binds to the molecule at the 9,10-position of the phenanthrenoid bond, which is known as the K -region, was taken as a model for the connection of cellular structures with polycyclic hydrocarbons. Pullman's model assumes that the carcinogenic activity of molecules is the higher, the greater the electron charge accumulates in the K -region. For some polycyclic hydrocarbons, a correlation was found between the K -region bond order and the rate of addition of osmium tetroxide. A new model is proposed here, an alternative to the A. and B. Pullman model [9].

incorporated into DNA. In the experiment of P.Brooks and P.D.Lawley it was established with high accuracy [12] that polycyclic aromatic hydrocarbons link to DNA, and the carcinogenic activity is proportional to the number of hydrocarbon molecules involved in complexation. The target of the action of carcinogens in the cell is mainly DNA and, possibly, some proteins that play a structural role in the organization of the genome or are involved in the replication, recombination, or repair of the damaged DNA structure. Table 1 presents the following additive contributions to the interaction energy: E_{elect} is electrostatic energy, E_{pol} is the energy of polarization interaction, E_{disp} is dispersion energy, and E_{rep} is the energy of short-range exchange repulsion. A statistical analysis of the relationship between the molecular descriptor Z and the additive contributions to the energy of the intermolecular pair interaction showed that there is a

statistically significant relationship between the molecular factor Z and the contribution to the energy of short-range interactions: repulsion, polarization contribution, and dispersion interaction. For the contribution E_{rep} , the following regression equation was obtained:

$$E_{\text{rep}}(Z) = a_0 + a_1 Z, \quad N = 11, \quad R = 0.89 \pm 0.15, \quad R^* = 0.90 > R_{0.05}^{\text{cr}}(N-2) = 0.602; \quad a_0 = -12.33 \pm 2.88, \quad a_1 = 5.83 \pm 1.01; \quad t(a_1) = 5.76 > |t(a_0)| = 4.28 > t_{0.05}^{\text{cr}}(N-2) = 2.262; \quad RMSE = 0.197; \quad \text{sample size sufficient for the validity of the correlation coefficient: } N_{0.05}^{\text{min}} = 5; \quad F = 30.3 > F_{0.05}^{\text{cr}}(f_1=1; f_2=9) = 5.1; \quad \text{straightforwardness: } K = (N(1-R^2))^{0.5} = 1.57 < K^{\text{thr}} = 3.00 \quad [15]. \quad (1)$$

For small sample sizes $N < 50$, the significance of the correlation coefficient is estimated using the Fisher normalizing z -transform, taking into account Hotelling's corrections [16]:

$$u_H = z - (3z + |R|)/4N = 1.303 > u_{0.05}(N) = z_{0.975} \cdot (N-1)^{-0.5} = 0.620, \quad z = 0.5 \cdot \ln((1+|R|)/(1-|R|)) = 1.42. \quad (2)$$

Here, for the quantile of the normal distribution, the value $z_\alpha = 1.960$ (the significance $\alpha = 0.975$); R^* is the corrected correlation coefficient [16]:

$$R^* = [1 + 0.5(1-R^2)/(N-3)], \quad (3)$$

which is used for small sample sizes $N \leq 15$ [17]. The value of the correlation coefficient R (1) falls within the range of 0.7-0.9, which, according to Chaddock's classification, is characterized as a close relationship.

Population statistics of the factor Z (in arb. units):

$$N = 11; \quad Z^{\text{av}} = 2.84 \pm 0.02; \quad 95\% \text{ confidence interval: } (2.80 - 2.88); \quad Z^{\text{min}} = 2.75, \quad Z^{\text{max}} = 2.94, \quad S_Z = 0.062, \quad \tau^{\text{min}} = 1.45 < \tau^{\text{max}} = 1.61 < \tau_{0.05}^{\text{cr},2}(N) = 2.343 < \tau_{0.05}^{\text{cr},1}(N) = 2.484; \quad \text{Wilk-Shapiro normality test: } W = 0.943 > W_{0.05}^{\text{cr}}(N) = 0.850, \quad \text{David-Hartley-Pearson normality test: } U1_{0.05}^{\text{cr}}(N) = 2.74 < U = [(Z^{\text{max}} - Z^{\text{min}})/S_Z] = 3.06 < U2_{0.05}^{\text{cr}}(N) = 3.80. \quad (4)$$

Statistics of the population E_{rep} :

$$N = 11; \quad E_{\text{rep}}^{\text{av}} = 4.25 \pm 0.12; \quad 95\% \text{ confidence interval: } (3.97 - 4.52); \quad E_{\text{rep}}^{\text{min}} = 3.67, \quad E_{\text{rep}}^{\text{max}} = 4.95, \quad S_{E_{\text{rep}}} = 0.405, \quad \tau^{\text{min}} = 1.43 < \tau^{\text{max}} = 1.73 < \tau_{0.05}^{\text{cr},2}(N) = 2.343 < \tau_{0.05}^{\text{cr},1}(N) = 2.484; \quad \text{Wilk-Shapiro}$$

$$\text{normality test: } W = 0.960 > W_{0.05}^{\text{cr}}(N) = 0.850, \quad \text{David-Hartley-Pearson normality test: } U1_{0.05}^{\text{cr}}(N) = 2.74 < U = [(E_{\text{rep}}^{\text{max}} - E_{\text{rep}}^{\text{min}})/S_{E_{\text{rep}}}] = 3.16 < U2_{0.05}^{\text{cr}}(N) = 3.80. \quad (5)$$

Thus, the elements of the sets Z and E_{rep} are homogeneous and normally distributed. The correlation coefficient is a reliable indicator of the relationship of features if the sample is homogeneous. The homogeneity of the samples was checked using the Grubbs - Romanovsky τ -test [17].

The relationship between the pair energies E_{pol} and E_{disp} with the molecular factor Z has the following statistics:

$$E_{\text{pol}}(Z) = a_0 + a_1 Z, \quad N = 11, \quad R = -0.76 \pm 0.22, \quad |R^*| = 0.78 > R_{0.05}^{\text{cr}}(N-2) = 0.602, \quad a_0 = 3.36 \pm 1.20, \quad a_1 = -1.48 \pm 0.42; \quad |t(a_1)| = 3.62 > |t(a_0)| = 2.81 > t_{0.05}^{\text{cr}}(N-2) = 2.262; \quad RMSE = 0.082; \quad \text{sample size sufficient for the validity of the correlation coefficient: } N_{0.05}^{\text{min}} = 7; \quad F = 12.4 > F_{0.05}^{\text{cr}}(f_1=1; f_2=9) = 5.1; \quad \text{straightforwardness: } K = (N(1-R^2))^{0.5} = 2.15 < K^{\text{thr}} = 3.00; \quad u_H = 1.045 > u_{0.05}(N) = 0.620; \quad (6)$$

$$E_{\text{disp}}(Z) = a_0 + a_1 Z, \quad N = 11, \quad R = -0.76 \pm 0.22, \quad |R^*| = 0.78 > R_{0.05}^{\text{cr}}(N-2) = 0.602, \quad a_0 = 26.26 \pm 9.89, \quad a_1 = -12.14 \pm 3.48; \quad |t(a_1)| = 3.49 > |t(a_0)| = 2.66 > t_{0.05}^{\text{cr}}(N-2) = 2.262; \quad RMSE = 0.678; \quad \text{sample size sufficient for the validity of the correlation coefficient: } N_{0.05}^{\text{min}} = 7; \quad F = 12.2 > F_{0.05}^{\text{cr}}(f_1=1; f_2=9) = 5.1; \quad \text{straightforwardness: } K = (N(1-R^2))^{0.5} = 2.15 < K^{\text{thr}} = 3.00; \quad u_H = 1.045 > u_{0.05}(N) = 0.620. \quad (7)$$

Statistics of the populations E_{pol} and E_{disp} :

$$N = 11; \quad E_{\text{pol}}^{\text{av}} = -0.85 \pm 0.04; \quad 95\% \text{ confidence interval: } (-0.93, -0.77); \quad E_{\text{pol}}^{\text{min}} = -1.11, \quad E_{\text{pol}}^{\text{max}} = -0.65, \quad S_{E_{\text{pol}}} = 0.120, \quad \tau^{\text{max}} = 1.67 < \tau^{\text{min}} = 2.17 < \tau_{0.05}^{\text{cr},2}(N) = 2.343 < \tau_{0.05}^{\text{cr},1}(N) = 2.484 \quad [17]; \quad \text{Wilk-Shapiro normality test: } W = 0.918 > W_{0.05}^{\text{cr}}(N) = 0.850, \quad \text{David-Hartley-Pearson normality test: } U1_{0.05}^{\text{cr}}(N) = 2.74 < U = [(E_{\text{pol}}^{\text{max}} - E_{\text{pol}}^{\text{min}})/S_{E_{\text{pol}}}] = 3.83 \approx U2_{0.05}^{\text{cr}}(N) = 3.80; \quad (8)$$

$N = 11$; $E_{\text{disp}}^{\text{av}} = -8.23 \pm 0.30$; 95% confidence interval: $(-8.90, -7.57)$; $E_{\text{disp}}^{\text{min}} = -10.37$, $E_{\text{disp}}^{\text{max}} = -6.84$, $S_{\text{Edisp}} = 0.986$, $\tau^{\text{max}} = 1.41 < \tau^{\text{min}} = 2.17$ $\tau_{0.05}^{\text{cr},2}(N) = 2.343 < \tau_{0.05}^{\text{cr},1}(N) = 2.484$; Wilk-Shapiro normality test: $W = 0.945 > W_{0.05}^{\text{cr}}(N) = 0.850$, David-Hartley-Pearson normality test: $U1_{0.05}^{\text{cr}}(N) = 2.74 < U = [(E_{\text{disp}}^{\text{max}} - E_{\text{disp}}^{\text{min}})/S_{\text{Edisp}}] = 3.58 < U2_{0.05}^{\text{cr}}(N) = 3.80$. (9)

Thus, all analyzed populations are homogeneous and have a distribution close to normal.

The electronic factor Z is statistically significantly associated with the short-range (E_{rep} , E_{pol} and E_{disp}) components of intermolecular interactions, which determine the formation of complexes of organic substances with molecular systems that simulate biosystems. At the same time, there is no statistically significant relationship between factor Z and the long-range electrostatic contribution E_{elect} . However, it is interesting to note the following detail. Quantum-mechanical analysis of the electrostatic energy of interaction between nucleotide bases [18] revealed convergence oscillations of the electrostatic contribution. If we arrange the molecules of Table 1 in descending order according to the Z factor, then the relationship between the electrostatic contribution and the value of the Z factor also has an oscillatory character. There is convincing evidence [19] of the binding of polycyclic hydrocarbons to natural nucleic acids, synthetic polynucleotides and their components, obtained by various methods: spectral (by the shift of the optical absorption maxima, the shift of the fluorescence maximum to the long-wavelength region), centrifugation, and the radiochemical method [20-30]. The degree of binding of polycyclic hydrocarbons with nucleic acids depends [19] on the conditions of complex formation, that is, the dispersion of the hydrocarbon particles, the pH of the medium, and ionic strength. The formation of complexes sharply decreases with increasing pH and ionic strength of the solution [31,32]. It is possible that these conditions can serve as limiting factors in the manifestation of carcinogenic activity by polycyclic aromatic hydrocarbons.

Statistical analysis of the relationship between molecular factors Z and H for polycyclic

hydrocarbons showed their close relationship. The linear relationship of factors Z and H has the following statistics:

$H(Z) = a_0 + a_1 \cdot Z$, $N = 11$, $R = -0.96 \pm 0.09$, $|R^*| = 0.965 > R_{0.05}^{\text{cr}}(N-2) = 0.602$; $a_0 = 1.49 \pm 0.05$, $a_1 = -0.186 \pm 0.02$, $RMSE = 0.0035$, $t(a_0) = 29.6 > |t(a_1)| = 10.48 > t_{0.05}^{\text{cr}}(N-2) = 2.26$, $F = 109.98 > F_{0.05}^{\text{cr}}(f_1=1, f=9) = 5.1$; $u_H = 2.014 > u_{0.05}(N) = 0.620$; straightforwardness: $K = (N(1 - R^2))^{0.5} = 0.93 < K^{\text{thr}} = 3.00$. (10)

Statistics of the set of information function:

$N = 11$; $H^{\text{av}} = 0.963 \pm 0.004$; 95% confidence interval: $(0.954 - 0.971)$; $H^{\text{min}} = 0.94$, $H^{\text{max}} = 0.98$, $S_H = 0.012$, $\tau^{\text{max}} = 1.42 < \tau^{\text{min}} = 1.92 < \tau_{0.05}^{\text{cr},2}(N) = 2.343 < \tau_{0.05}^{\text{cr},1}(N) = 2.484$; Wilk-Shapiro normality test: $W = 0.916 > W_{0.05}^{\text{cr}}(N) = 0.850$, David-Hartley-Pearson normality test: $U1_{0.05}^{\text{cr}}(N) = 2.74 < U = [(H^{\text{max}} - H^{\text{min}})/S_H] = 3.33 < U2_{0.05}^{\text{cr}}(N) = 3.80$. (11)

The set of elements Z and H are homogeneous and close to normal distribution. An increase in the sample size leads to the appearance of a weak nonlinearity (exponential) for the relationship of features, while the $RMSE$ decreases to a value of 0.0005 (Fig. 1). That is, the relationship approaches functional dependence.

Hereinafter, the observations that were used by the A. and B. Pullman to construct and substantiate their model are analyzed. Using the data of Pullman [9], we compose the following sequence of polycyclic hydrocarbons containing from 3 to 5 condensed rings:

Three rings: anthracene (–; $Z = 2.750$ arb. units; $H = 0.98$ bits), phenanthrene (–; $Z = 2.750$ arb. units; $H = 0.98$ bits); four rings: tetraphene (+/–; $Z = 2.800$ arb. units; $H = 0.97$ bits), benzo(c)phenanthrene (+/–; $Z = 2.800$ arb. units; $H = 0.97$ bits), pyrene (+/–; $Z = 2.846$ arb. units; $H = 0.96$ bits); five rings: dibenz-1,2,7,8-anthracene (+; $Z = 2.833$ arb. units; $H = 0.96$ bits), dibenz-1,2,3,4-phenanthrene (+; $Z = 2.833$ arb. units; $H = 0.96$ bits), dibenz-1,2,5,6-anthracene (++; $Z = 2.833$ arb. units), benz-3,4-pyrene (++++; $Z = 2.875$ arb. units; $H = 0.955$ bits). (12)

A

B

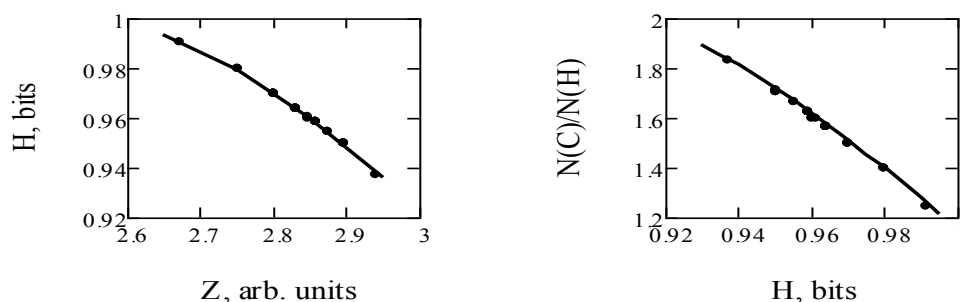


Figure 1. Cyclic aromatic hydrocarbons. A. Relationship between molecular factors H and Z . B. Relationship between molecular factor attributes $\gamma = N(C)/N(H)$ and H .

The level of carcinogenic activity (–, +/–, +,) and the values of the molecular factors Z and H are indicated in brackets. To quantitatively determine the relationship between bioactivity and the value of the factor Z , we use the Spearman's rank correlation coefficient [33]. In order to use this criterion, it is necessary to move from empirical values of features or qualitative characteristics to ranks. The rank refers to the series of a random variable in an ascending sample (ranking by carcinogenicity). The rank determines the position of the factor in the common list and is written as a natural number. The assessment of the relationship of factors is carried out not by the numerical values of the populations, but by the values of the ranks. Taking into account that there are connectives for the analyzed series of compounds [34], the rank correlation coefficient is determined from the relation:

$$\rho = \left[\frac{N(N^2 - 1)/6 - \sum_i (v_i - w_i)^2 - A - B}{[(N(N^2 - 1)/6 - 2A)(N(N^2 - 1)/6 - 2B)]^{0.5}} \right] \times$$

$$(13)$$

Here, A and B are corrections due to the connectives

$$A = \sum_i (A_i^2 - A_i), \quad (14)$$

$$B = \sum_j (B_j^2 - B_j). \quad (15)$$

A_i and B_j is the number of elements in the i and j -th connectives; subscript A refers to bioactivity and subscript B refers to explanatory variable.

$$\rho = 0.949 > \rho_{0.05}^{\text{cr}}(N = 9) = 0.5833 \quad [16]. \quad (16)$$

Since the calculated value exceeds the tabular value of the rank correlation (for sample sizes $N < 30$), inequality (16) indicates the presence of significant

positive parallelism of the effective and explanatory features. It is generally accepted that if ρ is in the

range of 0.7–0.99, then the statistical relationship is strong. Additionally, you can also use the following relationship, from which we get the usual correlation coefficient:

$$r = 2\sin(\pi\rho/6) = 0.953 > r_{0.05}^{\text{cr}}(N - 2) = 0.632. \quad (17)$$

The sequence of polycyclic compounds (11) was used by Pullman to substantiate his proposed model. It is obvious that an increase in the factor Z (or a decrease in the value of factor H) is accompanied by an increase in the carcinogenic activity of polycyclic hydrocarbons. That is, there is a positive parallelism between the level of carcinogenic activity of polycyclic aromatic hydrocarbons and the value of their molecular factor Z . In this regard, we note that two dibenzphenanthrene XIII and XIV have a value of $Z = 2.83$ arb. units (numbering of hydrocarbons according to Pullman [35]) do not contain the reaction K -region and cannot be carcinogenic. The Pullman model assumes that the K -region of molecules should have a high electron density for carcinogens, which is noticeably higher than the electron density for molecules that do not have carcinogenic activity. However, the use of factor Z makes it possible to differentiate the activity of hydrocarbons from the sequence (11) more accurately. So the value of the electronic factor $Z = 2.83$ arb. units allows for the possibility of low carcinogenic activity for these polycyclic hydrocarbons. This is confirmed by observations [36]. The absence of carcinogenic activity is characteristic of molecules containing less than four condensed benzene rings: benzene ($Z = 2.50$ arb. units, $H = 1.00$ bits), naphthalene ($Z = 2.67$ arb. units, $H = 0.99$ bits), anthracene ($Z = 2.75$ arb. units, $H = 0.98$ bits), phenanthrene ($Z = 2.75$ arb. units, $H = 0.98$ bits). For these hydrocarbons, the value of the factor Z is below 2.80 arb. units. The situation is similar for pentaphene, for which the K -region reactivity is close to the threshold value determined by the Pullman model. It follows from the Pullman model that this polycyclic hydrocarbon can have

only a very weak carcinogenic activity, which is usually denoted as follows: (+/-). The value of the electronic descriptor $Z = 2.83$ arb. units confirms this conclusion. The value of the electronic descriptor $Z = 2.83$ arb. units confirms this conclusion. Of more than fifty cyclic hydrocarbons, the Pullmans proposed to distinguish three groups of carcinogenic substances, which are arranged according to the level of their carcinogenic activity in the following sequence:

Group I: benz-3,4-phenanthrene ($Z = 2.80$ arb. units, $H = 0.97$ bits), dibenz-1,2,5,6-phenanthrene ($Z = 2.833$ arb. units, $H = 0.964$ bits), dibenz-1,2,3,4-phenanthrene ($Z = 2.83$ arb. units, $H = 0.964$ bits);

Group II: dibenz-1,2,5,6-anthracene ($Z = 2.833$ arb. units, $H = 0.964$ bits), dibenz-1,2,7,8-anthracene ($Z = 2.833$ arb. units, $H = 0.964$ bits);

Group III (the most active carcinogens): benz-3,4-pyrene ($Z = 2.875$ arb. units, $H = 0.955$ bits), dibenz-1,2,3,4-pyrene ($Z = 2.895$ arb. units, $H = 0.950$ bits), dibenz-3,4,8,9-pyrene ($Z = 2.895$ arb. units, $H = 0.950$ bits), dibenz-3,4,9,10-pyrene ($Z = 2.895$ arb. units, $H = 0.950$ bits). (18)

Thus, the use of the electronic factor Z (as well as factor H) allows us to make a comparative assessment of the levels of carcinogenicity for each group of polycyclic hydrocarbons from the sequence (18). It should be noted that the three groups of compounds identified by the Pullman retain their sequence according to the factor Z (and also according to the factor H). The discovered positive parallelism between the levels of carcinogenic activity of polycyclic hydrocarbons and the value of the molecular factor Z can be associated with the ability of these molecules to enter into complexation due to intermolecular interactions. The higher the Z value, the more likely the stronger the intermolecular bond with DNA and RNA (see equations (1), (6) and (7). For example, there are data [37, 38] indicating that the complexation of a benz-3,4-pyrene molecule with DNA and RNA depends on the ratio [adenine (A) + thymine (T)]/[guanine (G) + cytosine (C)] in DNA, and the G ($Z = 3.50$ arb. units) + C ($Z = 3.23$ arb. units) regions bind benz-3,4-pyrene more preferentially than the A ($Z = 3.33$ arb. units) + T ($Z = 3.20$ arb. units) regions. The total value of the factors Z for the G + C site is 6.73 arb. units, that is, noticeably higher than 6.50 arb. units for the A + T site. In addition, it should be noted that the values of factors Z and H for all purine and pyrimidine bases are significantly higher than for all analyzed polycyclic aromatic hydrocarbons. These results are very

important for understanding the possible specificity of the mutagenic action of chemical carcinogens. Depurination of DNA can lead to the formation of single-strand breaks, which can be the reason for a decrease in the activity of DNA repair and the subsequent development of carcinogenesis [19]. Apparently, the specificity of the mutagenic action of chemical carcinogens can be associated with the fact that they "distinguish" DNA regions containing an unequal percentage of G + C pairs [39].

The Pullman studies indicate that their proposed criteria for assessing the level of carcinogenicity for polycyclic aromatic hydrocarbons have exceptions. First, benz-3,4-phenanthrene, dibenz-1,2,5,6-phenanthrene and dibenz-1,2,3,4-phenanthrene do not obey the criteria established by them. At the same time, the use of the electronic factor Z (or factor H) does not negate the weak carcinogenicity [40] of these compounds. Secondly, according to observations, anthanthrene should have carcinogenic activity, although it does not have a highly active region K . This is also indicated by the value of the factor $Z = 2.94$ arb. units. According to modern data [36], anthanthrene does indeed have carcinogenic activity. Third, according to [38], the strong carcinogens dibenz-3,4,6,7-pyrene and dibenz-3,4,8,9-pyrene do not have an active K -region. Obviously, the carcinogenic activity of drugs assumed according to the Pullman's model cannot be high. At the same time, the molecular factor for these polycyclic hydrocarbons is very high ($Z = 2.895$ arb. units) and, therefore, indicates a strong carcinogenic activity of these hydrocarbons. This result does not contradict observations [36]. We note further the following that the most studied addition reactions occurring in the L region of the molecule are according to Pullman: a) fixation of maleic anhydride and b) photooxidation. It is assumed that the L -region of the molecule promotes the binding of the molecule in the *para*-position since it has the lowest values of the *para*-localization energy. Electronic factor Z (or information factor H) in this case also makes it possible to trace the degree of ease of fixation of maleic anhydride. The Pullmans pointed out that the ease of fixation of a molecule decreases in the following sequence:

Anthracene ($Z = 2.75$ arb. units, $H = 0.98$ bits; -) > benz-1,2-anthracene ($Z = 2.80$ arb. units, $H = 0.97$ bits; +/-) > dibenz-1,2,5,6-anthracene ($Z = 2.83$ arb. units, $H = 0.96$ bits; ++) > benz-3,4-pyrene ($Z = 2.875$ arb. units, $H = 0.96$ bits; ++++). (19)

This sequence of hydrocarbons was proposed by the Pullmans to validate their model. Photooxidation

very easily occurs in non-carcinogenic molecules of anthracene ($Z = 2.75$ arb. units) and naphthacene ($Z = 2.80$ arb. units) and becomes more difficult as additional side rings are attached. So, the molecule of dibenz-1,2,5,6-anthracene ($Z = 2.83$ arb. units) very difficultly enters into such a reaction, and the molecule of benz-3,4-pyrene ($Z = 2.875$ arb. units) already absolutely inactive in the photooxidation reaction. Obviously, the sequence of inequalities (19) is associated with the value of the factor Z . This gradation of ease of fixation is confirmed by experience. As the ease of fixation of maleic anhydride decreases, an increase in the carcinogenic activity of hydrocarbons and a simultaneous increase in the Z factor is observed. In accordance with the sequences (18) and (19), it is possible to approximately indicate the range of values of the factor attribute $Z > Z^{\text{thr}} \approx 2.80 - 2.82$ arb. units, which can be taken as the range of threshold values. This interval separates carcinogenic polycyclic aromatic hydrocarbons from non-carcinogenic ones. A number of polycyclic compounds proposed by Pullmans (19) can be expanded by adding, for example, non-carcinogenic tetracene (gross formula $C_{18}H_{12}$; $Z = 2.80$ arb. units; $H = 0.971$ bits). Methylation of the hydrocarbon increases the reactivity of the chemical compound. Thus, 9,10-dimethylbenzanthracene ($Z = 2.67$ arb. units, $H = 0.99$ bits) is much more reactive than the starting hydrocarbon benzanthracene ($Z = 2.80$ arb. units, $H = 0.97$ bits) in reactions with maleic anhydride and in photooxidation reactions. All these experimental facts are also interconnected with the dynamics of changes in the electronic factor Z (or H). The following polycyclic compounds have high carcinogenic activity [41]: dibenz-1,2,3,4-pyrene ($Z = 2.895$ arb. units; $H = 0.950$ bits), dibenz-3,4,8,9-pyrene ($Z = 2.895$ arb. units; $H = 0.950$ bits), dibenz-3,4,9,10-pyrene ($Z = 2.895$ arb. units; $H = 0.950$ bits). You can also add to this series of hydrocarbons, for example, the ovalene molecule ($Z = 3.087$ arb. units; $H = 0.887$ bits). The ovalene molecule has pronounced carcinogenic, mutagenic and teratogenic properties. As was shown in [42], dibenz-3,4,6,7-pyrene is also very active (the blastomogenic substance is no less active than benz-3,4-pyrene). For this chemical compound, the descriptors $Z = 2.895$ arb. units and $H = 0.950$ bits. The Pullmans applied their model to predicting the carcinogenic activity of dibenz-3,4,6,7-pyrene. Their calculations also point to the carcinogenicity of this hydrocarbon. The Pullmans model assumes that the presence of the L -region in the molecule of polycyclic hydrocarbons does not contribute to the appearance of the reactive activity of the molecule

[35]. Indeed, according to the decreasing ability to react with $Pb(OAc)_4$, the hydrocarbons are arranged in the following sequence (this row was proposed by the Pullmans):

Anthracene ($Z = 2.75$ arb. units; $H = 0.98$ bits; $-$) $>$ benz-1,2-anthracene ($Z = 2.80$ arb. units; $H = 0.97$ bits; $+/-$) $>$ dibenz-1,2,5,6-anthracene ($Z = 2.83$ arb. units, $H = 0.96$ bits; $+$) $= 0$. (20)

The sequence of carcinogenicity (20) was used by the Pullmans to confirm the adequacy of their model to the observed facts. The application of the molecular explanatory factors Z and H in this case also does not contradict these inequalities.

To study the reactivity of the K -region of polycyclic hydrocarbons, the reaction of addition of osmium tetroxide was studied in detail. Quantitative studies of the reactivity of the K -region allowed G.M. Bedger [43] to compose the following sequence of chemical compounds according to their increasing ability to react with osmium tetroxide:

Phenanthrene ($Z = 2.75$ arb. units; $I = 7.75$ eV; $-$) $<$ benz-1,2-anthracene ($Z = 2.80$ arb. units; $I = 7.50$ eV; $+/-$) $<$ dibenz-1,2, 5,6-anthracene ($Z = 2.83$ arb. units; $I = 7.09-7.80$ eV; $++$) $<$ benz-3,4-pyrene ($Z = 2.875$ arb. units; $I = 7.19$ eV; $++++$). (21)

The series of polycyclic hydrocarbons (21) are arranged in order of increasing their reactivity with respect to the reaction with osmium tetroxide. The sequence of inequalities (21) demonstrates that in this case there is a parallelism of reactivity with the value of the molecular factor sign Z . The relative carcinogenic activity of molecules increases with increasing Z value. Inequalities (21) also give the values of the first ionization potentials I of polycyclic aromatic hydrocarbons, which were determined from the electronic spectra with charge transfer [44]. In (21) according to Pullman [9], the carcinogenic activity of chemical compounds is given. As a result, of the conducted studies, it was found that an increase in carcinogenicity and factor Z of polycyclic hydrocarbons is accompanied by a tendency for a decrease in the value of the ionization potential of the molecule. It is known, the lower the ionization energy of the molecule, the better the donor properties of the hydrocarbon. An increase in the carcinogenicity and factor Z of molecules is often associated with an increase in the donor capacity of the molecules. There is a relationship between the value of the factor attribute Z and the value of the energy of the first ionization potential I (21) of polycyclic aromatic hydrocarbons. For a limited amount of data (20), the following linear regression equation was obtained:

$$I(Z) = a_1 + b_1 \cdot Z, N = 4, R = -0.988 \pm 0.107, |R^*| = 0.999 > R_{0.05}^{cr}(N-2) = 0.950; a_1 = 19.63 \pm 1.32, b_1 = -4.32 \pm 0.47, t(a_1) = 14.9 > |t(b_1)| = 9.24 > t_{0.05}^{cr}(N-2) = 2.306; RMSE = 0.043; F = 85.39 > F_{0.05}^{cr}(f_1=1; f_2=2) = 18.51; \text{straightforwardness: } K = (N \cdot (1 - R^2))^{0.5} = 0.22 < K^{thr} = 3.00. \quad (22)$$

Here the average value of the ionization potential is taken for the dibenz-1,2,5,6-anthracene molecule. Statistics of the sets Z and I :

$$N_1 = 4, Z^{cp} = 2.81 \pm 0.03; 95\% \text{ confidence interval: } 2.73 - 2.90; Z^{\min} = 2.75, Z^{\max} = 2.875, S_Z = 0.053, \tau^{\min} = 1.13 < \tau^{\max} = 1.23 < \tau_{0.05}^{cr,2}(N) = 1.889 < \tau_{0.05}^{cr,1}(N) = 1.941; \text{Wilk-Shapiro normality test: } W = 0.999 > W_{0.05}^{cr}(N) = 0.748;$$

$$N_1 = 4, I^{cp} = 7.47 \pm 0.12; 95\% \text{ confidence interval: } 7.11 - 7.84; I^{\min} = 7.19, I^{\max} = 7.75, S_I = 0.23, \tau^{\max} = 1.21 < \tau^{\min} = 1.23 < \tau_{0.05}^{cr,2}(N) = 1.889 < \tau_{0.05}^{cr,1}(N) = 1.941; \text{Wilk-Shapiro normality test: } W = 0.978 > W_{0.05}^{cr}(N) = 0.748. \quad (23)$$

For the purpose of express analysis of the carcinogenic activity of angularly annelated aromatic hydrocarbons, one more additional molecular factor can be introduced, which is determined by the ratio of the number of carbon atoms $N(C)$ in a polycyclic molecule to the number of hydrogen atoms $N(H)$: $\gamma = N(C)/N(H)$. This molecular factor can be called a structural factor. As shown by statistical analysis, factor γ , determined independently of the molecular factors Z and H , correlates with them. For thirty-six polycyclic aromatic hydrocarbons presented in the book by B. Pullman [9], there is indeed a statistically significant close relationship between the characteristics. For example, the relationship between information function H and structural factor γ can be approximated by the following equation, which has a very low $RMSE$:

$$\gamma(H) = a_0 + a_1 \cdot \exp(-(H - a_2)^2/a_3^2), N = 36, RMSE = 0.005, a_0 = -3.73 \pm 14.3, a_1 = 5.94 \pm 14.6, a_2 = 0.85 \pm 0.07, a_3 = -0.33 \pm 0.55. \quad (24)$$

There is also a close relationship between the factors Z and γ (Fig. 2). Taking into account the area of boundary values for the factor attribute Z , as well as the functional dependence $\gamma(Z)$, it is possible to approximately indicate the area of boundary values for the structure factor: $\gamma > \gamma^{thr} \approx 1.50-1.55$.

Polycyclic aromatic hydrocarbons having γ values that exceed this region are likely to be potentially active carcinogens.

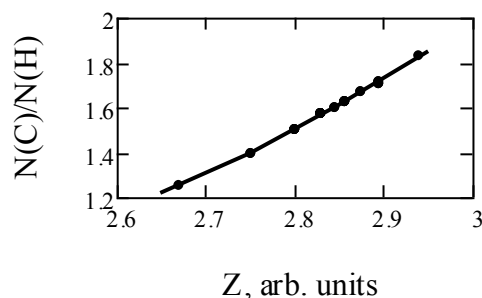


Figure 2. Relationship between molecular factors Z and γ . Factor relationship equation: $\gamma(Z) = A + B \cdot \exp(-C \cdot Z)$, $N = 36$, $A = -0.208 \pm 0.152$, $B = 0.0547 \pm 0.0213$, $C = -1.230 \pm 0.107$, $RMSE = 0.003$.

The statistically significant relationship between the molecular factors Z , H , and γ indicates that these factors are equal for the interpretation of the carcinogenicity of polycyclic aromatic hydrocarbons and the accompanying phenomena. It is important to note that the molecular factors that characterize the molecule as a whole are obtained on the basis of different principles, are closely interrelated with each other and lead, in the analysis of the carcinogenic activity of molecules, for angularly annelated aromatic hydrocarbons to identical results.

An increase in the sample size by including polycyclic hydrocarbons containing more than six benzene rings in the sample preserves the relationship of factors. In quantitative terms, this relationship leads to the appearance of a weak nonlinearity between the factors γ and H . Since the factors Z , H and γ for polycyclic aromatic hydrocarbons are correlated with each other the quantitative assessment of the level of carcinogenicity can be performed using the easily calculated structural factor γ . The above observed facts [9,35,40] for polycyclic aromatic hydrocarbons can be interpreted using an additional structural factor-attribute γ . For example, the level of carcinogenic activity of chemical compounds increases in the following sequence:

$$\text{Phenanthrene } (\gamma = 1.40; -) < \text{benz-1,2-anthracene } (\gamma = 1.50; +/ -) < \text{dibenz-1,2,5,6-anthracene } (\gamma = 1.57; ++/++) < \text{benz-3,4-pyrene } (\gamma = 1.67; ++++) \approx \text{dibenz-3,4,6,7-pyrene } (\gamma = 1.714; ++/++++) \approx \text{dibenz-3,4,8,9-pyrene } (\gamma = 1.714; ++++). \quad (25)$$

The series (25) can be extended by adding the ovalene molecule ($\gamma = 2.286$), which is a very strong

carcinogen. Obviously, the molecular factor γ is associated with the level of carcinogenic activity: an increase in the carcinogenicity of a substance is accompanied by an increase in the value of γ (as well as for the sign Z).

To quantitatively determine the relationship between bioactivity and the value of the sign Z , we use the Spearman's rank correlation coefficient [33, 17]. In the event that the sample does not contain a connectives, then the rank correlation coefficient is determined from the following relation [16]:

$$\rho = 1 - 6 \cdot \sum_{i=1}^N d_i^2 / [N(N^2 - 1)] = 0.843 > \rho_{0.05}^{\text{cr}}(N = 6) = 0.7714. \quad (26)$$

Here $d_i = v_i - w_i$ is the difference of ranks for each i -th pair of $N = 6$ observations. Active $++/+++$ was assigned a rank of 2.5, and an activity $+/++++$ was assigned a rank of 3. Inequality (26) indicates the presence of significant positive parallelism of the effective and explanatory features. The availability of calculations of molecular factors γ , Z and H allows them to be used for the purpose of express assessment of the carcinogenic activity of polycyclic aromatic hydrocarbons. As you know, the value of any model is determined by the combination of two of its qualities – the accuracy and availability of the solution to the problem under study.

Let us check how the established relationships change when the effect of hydrocarbon methylation is taken into account. So, for example, carcinogenic benz-3,4-pyrene has the following values of factors $\gamma = 1.67$, $Z = 2.875$ arb. units. The addition of methyl groups to benz-3,4-pyrene leads to a decrease in the carcinogenic activity of molecules and, at the same time, to a decrease in the values of molecular factors. Indeed, inactive 2'- and 3'-methyl derivatives have the following values of factor-signs $\gamma = 1.50$, $Z = 2.80$ arb. units. Methylation of benz-3,4-phenanthrene ($\gamma = 1.50$, $Z = 2.80$ arb. units) and chrysene ($\gamma = 1.50$, $Z = 2.80$ arb. units) also reduces the carcinogenic activity of 6,7-dimethyl-3,4-benzphenanthrene ($\gamma = 1.25$, $Z = 2.667$ arb. units) and 4,5-dimethylchrysene ($\gamma = 1.25$, $Z = 2.667$ arb. units). This decrease in bioactivity is accompanied by a decrease in the values of factors, which is in accordance with inequalities (12) and (19). From the point of view of the Pullman model, the absence of carcinogenic activity in these hydrocarbons is "... an extremely mysterious fact... and it is difficult to find any explanation for this curious phenomenon" [35]. However, methylation can increase carcinogenicity by reducing the value of the first ionization potential of molecules. At the same time,

in accordance with the sequence (21), this can reduce the carcinogenic activity by decreasing the factor Z , which in turn reduces the ability of the molecule to bind in the biosystem. The addition of a methyl group to a cyclic molecule usually leads to a bathochromic shift (to longer wavelengths) of the absorption maximum. This shift is accompanied by an increase in the π electronic level of the highest filled MO and, at the same time, a decrease in the lowest level of a free MO [45]. As a result, the ionization potential decreases. The addition of a radical (methyl group) to a cyclic compound leads to the so-called M – R conjugation. The influence of the substituent on the total π electronic system is carried out by the mesomeric M – effect.

A decrease in the molecular factor Z reduces the ability of the molecule to participate in complexation. At the same time, decreasing the ionization potential improves the donor properties of the molecule. That is, there is a competition between two factors. However, depending on the position of the methyl group in the ring, the direction of the level shifts can be reversed. The bathochromic shift can change to a hypsochromic shift (shift towards shorter wavelengths). This change was found in azulenes [9]. In this case, the ionization potential of the molecule increases. That is, in this situation, both factors have the same direction of influence on the level of carcinogenic activity of a substance - its decrease. A decrease in the carcinogenic activity of the parent molecule may also be associated with a decrease in the energy of the molecule's electron affinity (see below). It is known [44] that methylation often leads to a decrease in the energy of affinity of the molecule.

The threshold values of molecular factors for unsubstituted polycyclic aromatic hydrocarbons and their derivatives appear to be different. For polycyclic aromatic hydrocarbons, carcinogenic activity increases with an increase in molecular descriptors γ and Z . A decrease in the molecular characteristics γ and Z for methylated polycyclic hydrocarbons is accompanied by an increase in the reactivity of the substance in the reaction with lead tetraacetate. For example, 20-methylcholontrene ($\gamma = 1.31$, $Z = 2.70$ arb. units) and 20-methyl-1,2,3,4,11,14-hexahydrocholontrene ($\gamma = 0.99$, $Z = 2.47$ arb. units) have low values of γ and Z and are very reactive in the reaction with lead tetraacetate [46]. It is fundamentally important for this study that the concept of specific reaction regions of molecules (K and L regions) is not used here to explain the carcinogenic properties of polycyclic aromatic compounds. On the contrary, the model proposed here uses molecular factors that

characterize the properties of molecules as a whole. Obviously, within the framework of the new model, the following important remark of Ya. Ladik: "... why only hydrocarbons that have joined the *K*-region are carcinogenic, while those that have joined in the *L* position are inactive" [47]. It is important to note that the molecular factors *Z*, *H* and γ allow one to fully interpret all the observations that are fundamental to the Pullman model. Firstly, factors *Z*, *H* and γ are interrelated for polycyclic aromatic hydrocarbons, secondly, they are obtained on the basis of completely different properties of molecules and, thirdly, they all characterize the molecule as a whole. In accordance with the Pullman model, such polycyclic compounds as naphthalene ($Z = 2.67$ arb.units, $\gamma = 1.25$, $H = 0.991$ bits), anthracene ($Z = 2.75$ arb.units, $\gamma = 1.40$, $H = 0.980$ bits), tetracene ($Z = 2.80$ arb.units, $\gamma = 1.50$, $H = 0.971$ bits) and triphenylene ($Z = 2.80$ arb. units, $\gamma = 1.50$, $H = 0.971$ bits) do not have carcinogenic activity, since they lack a sufficiently active *K*-region [40]. The molecular factor *Z* for these substances also has low values, which are in the region below the boundary value Z^{thr} . This remark also applies to the boundary values $H^{\text{thr}} \approx 0.969$ bits (10) and $\gamma^{\text{thr}} \approx 1.50$ -1.55. For extremely active (+++, ++++) polycyclic aromatic hydrocarbons, the molecular factor *Z* is significantly higher than the threshold value and lies in a relatively narrow range of values:

Benz-3,4-pyrene ($Z = 2.875$ arb. units, $H = 0.955$ bits, $\gamma = 1.667$), dibenz-3,4,6,7-pyrene ($Z = 2.895$ arb. units, $H = 0.950$ bits, $\gamma = 1.714$), benzo-3,4-fluoranthene ($Z = 2.875$ arb. units, $H = 0.955$ bits, $\gamma = 1.667$), benzo-8,9-fluoranthene ($Z = 2.875$ arb. units, $H = 0.955$ bits, $\gamma = 1.667$), benzo(f)fluoranthene ($Z = 2.875$ arb. units, $H = 0.955$ bits, $\gamma = 1.667$), dibenz-3,4,8,9-pyrene ($Z = 2.895$ arb. units, $H = 0.950$ bits, $\gamma = 1.714$), dibenz-3,4,6,7-pyrene ($Z = 2.895$ arb. units, $H = 0.950$ bits, $\gamma = 1.714$), dibenz-3,4,9,10-pyrene ($Z = 2.895$ arb. units, $H = 0.950$ bits, $\gamma = 1.714$), dibenz-1,2,3,4-chrysene ($Z = 2.895$ arb. units, $H = 0.950$ bits, $\gamma = 1.714$), indeno[1,2,3-c,d]pyrene ($Z = 2.875$ arb. units, $H = 0.955$ bits, $\gamma = 1.667$).

(27)

According to A. and B. Pullman, dibenz-1,2,5,6-anthracene has a reduced activity (++) in comparison with a number of compounds (27). Indeed, for this chemical compound, the molecular characteristic *Z* decreases to 2.833 arb. units. According to Pullman's model, quantitative signs of carcinogenicity of polycyclic hydrocarbons must simultaneously satisfy two requirements [40] for the chemical activity of the *K*-region ($\leq 3.31\beta$) and *L*-

region ($\geq 5.66\beta$). β is the resonance integral of the interaction between orbitals within one chemical bond of valence-bound atoms in the molecular-orbital theory in the Hückel approximation. Using various experimental data, it was found [48] that the numerical value of the integral β for polycyclic hydrocarbons is in the range of values: - 0.69, - 2.71 eV. It follows from Pullman's model that the dibenz-3,4,6,7-pyrene molecule is presumably not carcinogenic [35,40]. However, a series of hydrocarbons (27) indicate that this polycyclic hydrocarbon ($Z = 2.895$ arb. units, $H = 0.95$ bits, electron affinity energy $A = 0.82$ eV [44]) belongs to strong carcinogens. This is confirmed by observations [42]. For our purposes, along with the *Z* factor, an important electronic property of a molecule is its affinity for an electron, which determines the amount of energy released or absorbed when an electron is attached to a molecule in its ground state. Unlike the ionization potential, which is always characterized by endoenergetic processes, the electron affinity is associated with both exoenergetic and endoenergetic processes. We also note that molecules with a high electron affinity energy value are strong oxidizing agents. Molecular factors *Z*, *H* and γ do not allow to distinguish between isomer molecules, for example, benz-3,4-pyrene (or benz-1,2-pyrene) ($Z = 2.875$ arb. units; very active) and benz-1,2-pyrene ($Z = 2.875$ arb. units; weakly active [36]. Both compounds are classified as alternate hydrocarbons. For both chemical compounds, the molecular factors are the same. Therefore, the conditions of complexation for them must be the same. Indeed, the data [19] on the binding constants with various polynucleotides indicate that the complexation of these compounds is approximately the same. The solubilities of benz-3,4-pyrene and benz-1,2-pyrene in aqueous solution are also close [27, 49]. However, it is not enough to establish the ability of a molecule to participate in complex formation. It is important to point out the possible processes that follow from this complexation and what consequences this can lead to the biosystem. However, such an important molecular characteristic as the electron affinity energy (*A*, eV) for these molecules differs markedly. Molecules with a high electron affinity are strong oxidizing agents. For benz-1,2-pyrene, the affinity energy, according to different literature sources, lies in the range of values $A = 0.07$ -0.4 eV. At the same time, for benz-3,4-pyrene, the affinity energy is noticeably higher and is equal to: $A = 0.77$ eV [44]. If for the affinity energy of the benz-1,2-pyrene molecule we take the lower limit of the

permissible range of values, then during the transfer of an electron in the complexation region with the participation of benz-3,4-pyrene, ten times more energy is released than for benz-1,2-pyrene. The released energy can be directly directed to the destruction of the ordered molecular structures of the biosystem. A similar situation takes place for carcinogenically active dibenz-1,2,5,6-anthracene (+++/++) and inactive dibenz-1,2,3,4-anthracene (–), the electronic and information factor signs of which match. At the same time, the values of their affinity energies differ significantly and are equal to 0.68 and 0.22 eV, respectively [44]. This difference in the value of the affinity limits the chemical activity of dibenz-1,2,3,4-anthracene. The low value of the energy of affinity can be a limiting factor for the manifestation of a clearly expressed carcinogenic activity by the coronene molecule, for which the factors Z and H are equal to 3.00 arb. units and 0.92 *bits*, respectively. Note the relatively high value of the first ionization potential of these molecules: 7.80 eV and 7.61 eV [44] which can also reduce the carcinogenic activity of hydrocarbons. For comparison, we present the approximate values of the affinity energy for purine (ionization potential $I \approx 8.3$ -8.9 eV) and pyrimidine bases (ionization potential $I \approx 8.0$ -8.3 eV): adenine ($A = -0.77$ eV), guanine ($A = -0.96$ eV), cytosine ($A = (-0.24, -1.25)$ eV) and uracil ($A = (+0.02, -1.13)$ eV) [44]. A negative value of the electron affinity means that the accession of an electron requires energy.

As indicated above, the electron affinity is a measure of the oxidative capacity of a molecule. The greater the electron affinity, the higher the oxidizing power of the substance. The oxidation process is accompanied by the transfer of an electron from a donor to an acceptor. Oxidation of a molecule can be accompanied by the appearance of instability in the molecular structure of a substance and thereby lead to its disintegration. The position on the energy scale of the level of the highest filled molecular orbital of benz-3,4-pyrene (the first ionization potential is 7.19 eV [44]) is noticeably higher compared to the deeper-lying highest filled energy level of benz-1,2-pyrene (the ionization potential 8.18 eV). An increase in the energy of affinity for molecules is accompanied by an increase in the carcinogenic activity of polynucleotides. (21). Consequently, the processes of electron transfer for benzo-3,4-pyrene are preferable. This is confirmed experimentally. A similar situation takes place for the dibenz-3,4,6,7-pyrene molecule ($Z = 2.895$ arb. units, $H = 0.95$ *bits*, $\gamma = 1.714$), which is characterized by a very high carcinogenic activity (level of activity ++/++++ [40]), and the molecule

also has a high affinity energy $A = 0.82$ eV. At the same time, for the isomer, that is, for the dibenz-3,4,8,9-pyrene molecule with a high value of the factor Z , there is no confirmed carcinogenic activity. For this molecule, the electron affinity energy is relatively low, $A = 0.5$ eV [44]. That is, the molecule can participate in complexation, since the factor Z is greater than the threshold value, however, the effect of this molecule on the biosystem is likely to be much less than that of the dibenz-3,4,6,7-pyrene molecule. A similar situation arises, for example, for coronene ($Z = 3.00$ arb. units; $H = 0.918$ *bits*; $\gamma = 2.00$; +/-). However, the affinity energy of coronene is very low: $A = 0.15$ – 0.3 eV [44]. The conjunction between the value of the energy of affinity and the carcinogenic activity of polycyclic hydrocarbons can be traced for the following series of compounds for which the molecular factor Z is large enough:

Dibenz-3,4,9,10-pyrene ($Z = 2.895$ arb. units, $H = 0.95$ *bits*, $A = 0.80$ - 0.84 eV [44]; ++++), dibenz-3,4,8,9-pyrene ($Z = 2.895$ arb. units, $H = 0.95$ *bits*, $A = 1.1$ eV [44]; ++++), dibenz-1,2,3,4-pyrene ($Z = 2.895$ arb. units, $H = 0.95$ *bits*, $A = 0.80$ - 0.87 eV [44]; ++++). (28)

All chemical compounds (28) are characterized by high values of both the affinity energy and the Z factor, which significantly exceeds its threshold value. That is, the participation of hydrocarbons in the formation of complexes, for example, with DNA and RNA, is very likely, followed by the impact on the biosystem of energy processes associated with the release of energy during electron transfer. At the same time, such polycyclic (six benzene rings) compounds similar in molecular structure as dibenz-1,2,6,7-pyrene ($Z = 2.895$ arb. units, $H = 0.95$ *bits*, $A = 0.05$ - 0.4 eV [44]) or tribenz-1,2,3,4,5,6-anthracene ($Z = 2.857$ arb. units, $H = 0.959$ *bits*, $A = 0.5$ eV [44]) have no carcinogenic activity. Apparently, in this case, the potential carcinogenicity of molecules (a large value of the factor Z) is limited by low values of the electron affinity.

A. Pullman and B. Pullman [35] to test the capabilities of their model identified the following series of polycyclic hydrocarbons containing five benzene rings. The carcinogenic activity of these molecules is significantly altered by small structural variations in the molecules. The sequence of carcinogenic activity for these hydrocarbons is as follows:

Benz-3,4-pyrene ($Z = 2.875$ arb. units; $H = 0.955$ *bits*; $\gamma = 1.667$; ++++; $A = 0.77$ eV) > dibenz-

1,2,5,6-anthracene ($Z = 2.833$ arb. units; ++; $H = 0.964$ bits; $\gamma = 1.571$; $A = 0.68$ eV) > dibenz-1,2,5,6-phenanthrene ($Z = 2.833$ arb. units; $H = 0.964$ bits; $\gamma = 1.571$; +; $A = 0.31$ eV) \approx dibenz-1,2,3,4-phenanthrene ($Z = 2.833$ arb. units; $H = 0.964$ bits; $\gamma = 1.571$; +; $A = 0.31$ eV) > dibenz-1,2,7,8-anthracene ($Z = 2.833$ arb. units; $H = 0.964$ bits; $\gamma = 1.571$; +; $A = 0.23$ eV) > benz-1,2-pyrene ($Z = 2.875$ arb. units; $H = 0.955$ bits; $\gamma = 1.667$; -; $A = 0.49$ eV) \approx dibenz-1,2,3,4-anthracene ($Z = 2.833$ arb. units; $H = 0.964$ bits; $\gamma = 1.571$; -; $A = 0.22$ eV);

$$\rho = 0.964 > \rho_{0.05}^{\text{cr}}(N) = 0.6786, r = 2\sin(\pi\rho/6) = 0.967 > r_{0.05}^{\text{cr}}(N-2) = 0.669. \quad (29)$$

Spearman's correlation coefficient was estimated using the formula (26). The ranking of chemical compounds was carried out on the basis of A , that is, the energy of affinity for an electron. The rank correlation coefficient ρ establishes the relationship between the electron affinity and carcinogenic activity. Thus, for the sequence (29) of polycyclic hydrocarbons, the electron affinity of the molecule also turned out to be informative. The Z factor for this series of hydrocarbons is above the limit value and, therefore, these compounds can take part in complexation. However, a decrease in electron affinity reduces the carcinogenic activity for molecules, even those with a sufficiently high value of the factor sign Z . That is, the magnitude of the affinity of a molecule for an electron can be a limiting factor for carcinogenic activity. The change in the carcinogenic activity of polycyclic compounds (29) depending on the electron affinity A (eV) is very similar to the threshold dependence. An approximate estimate of the possible threshold action can be obtained using the Cochran test [16]. Let's divide the sample (29) into two groups. The first group is active carcinogens (they have ++++ and ++ bioactivities). The second group contains compounds that have little bioactivity (+) or have no carcinogenic activity at all (-). Each bioactivity is assigned a number of points. For example, for activity (++++), four points are set, and for activity (-) zero points. According to the Cochran test, we obtain the following inequality:

$$\chi^2 = (N_1 \cdot N_2 / N) \cdot (A_1^{\text{av}} - A_2^{\text{av}})^2 / A^{\text{av}} = 6.40 > \chi_{0.05}^2(f=1) = 3.841. \quad (30)$$

Here $A_1^{\text{av}} = 3$ points, $A_2^{\text{av}} = 0.6$ points; overall average $A^{\text{av}} = 1.286$ points; $N_1 = 2$, $N_2 = 5$, $N = 7$. Since χ^2 is larger than the tabular *chi-square* value at the significance level $\alpha = 0.05$ and the number of degrees of freedom $f = 1$, it is possible that there is a

jump in carcinogenic activity approximately in the region of the affinity value of ~ 0.5 eV.

Let us also compare the factorial features for coronene (inactive or weakly active) and anthanthrene (active [36]) molecules. Both molecules have high values of the Z factor equal to 3.0 and 2.94 arb. units, respectively. Having a compact configuration and a high value of the Z factor, these molecules are able to participate in intermolecular complexation. However, the result of this interaction can vary significantly. Indeed, for coronene, the ionization potential and the affinity energy are $I = 7.50$ eV and $A = 0.30$ eV, respectively, while for anthanthrene the ionization potential is lower than $I = 7.10$ eV, and the affinity energy $A = 1.00$ eV, that is, more than three times higher. Apparently, in accordance with inequalities (21) and (29), the donor-acceptor properties of coronene can limit its potential carcinogenic activity.

Let us also check the relationship between the level of carcinogenicity of the hydrocarbon and the wavelength λ of the first absorption band due to charge transfer. For benzo-3,4-pyrene and benz-1,2-pyrene participating in the formation of the polycyclic hydrocarbon – tetracyanoethylene complex, the wavelengths are $\lambda = 820$ nm (1.51 eV) and 667 nm (1.86 eV), respectively [45]. The carcinogenic isomer is characterized by a longer wavelength λ . Possible mechanisms for limiting the activity of coronene have already been indicated above. At the same time, for this hydrocarbon, the wavelength of the absorption band due to charge transfer is small compared to benzo-3,4-pyrene and is equal to 725 nm [45]. Thus, according to this trait, coronene cannot be carcinogenic active substances, although it has a high value of the factor Z . By the size of the factors, perylene can be a potential carcinogen: $Z = 2.875$ arb. units, $\lambda = 0.920$ nm, $A = 0.85$ eV, $I = 6.80$ eV. Indeed, the possible carcinogenic activity of perylene is indicated in the handbook [36]. For the boundary value of the characteristic λ , it is approximately possible to take the value of the wavelength $\lambda^{\text{thr}} \approx 750$ nm. The limitation of carcinogenic activity is not noted above this value.

The limiting effect of the magnitude of the electron affinity energy of molecules can be traced using the following series of hydrocarbons proposed by Pullman. B. Pullman identified [9] the following polycyclic hydrocarbons, the carcinogenic activity of which changes significantly with small structural changes in the molecules:

Benz-3,4-pyrene ($Z = 2.875$ arb. units; $H = 0.955$

bits; $\gamma = 1.667$; $A = 0.77$ eV; +++) \approx dibenz-3,4,6,7-pyrene ($Z = 2.895$ arb. units; $H = 0.950$ *bits*; $\gamma = 1.714$; $A = 0.82$ eV; +++) $>$ dibenz-1,2,5,6-anthracene ($Z = 2.833$ arb. units; $H = 0.964$ *bits*; $\gamma = 1.571$; $A = 0.68$ eV; ++) $>$ dibenz-1,2,5,6-phenanthrene ($Z = 2.833$ arb. units; $H = 0.964$ *bits*; $\gamma = 1.571$; $A = 0.31$ eV; +) \approx dibenz-1,2,3,4-phenanthrene ($Z = 2.833$ arb. units; $H = 0.964$ *bits*; $\gamma = 1.571$; $A = 0.40$ - 0.43 eV; +) $>$ dibenz-1,2,7,8-anthracene ($Z = 2.833$ arb. units; $H = 0.964$ *bits*; $\gamma = 1.571$; $A = 0.23$ - 0.69 eV; +/-) $>$ benz-1,2-pyrene ($Z = 2.875$ arb. units; $H = 0.955$ *bits*; $\gamma = 1.667$; $A = 0.07$ - 0.40 eV; $H = 0.955$ *bits*; $\gamma = 1.667$; -) $>$ dibenz-1,2,3,4-anthracene ($Z = 2.833$ arb. units; $H = 0.964$ *bits*; $\gamma = 1.571$; $A = 0.22$ - 0.54 eV; -).

(31)

Experimental values of the electron affinity energy are reported in the handbook [44]. The lower bounds of the energy of affinity indicate their parallelism with the carcinogenic activity of molecules. The presence of a trend in the values of electron affinity can be checked using the Abbe-Linnik test [16, 17]:

$$q = 0.5 \cdot \sum_{i=1}^{N-1} (A_{i+1} - A_i)^2 / \sum_{i=1}^N (A_i - A^{\text{av}})^2 = 0.340 <$$

$$q_{0.05}^{\text{cr}}(N = 10) = 0.4912. \quad (32)$$

Here, the average value of the sample is $A^{\text{av}} = 0.507$ eV. Since inequality (32) holds, the null hypothesis about the randomness of the series should be rejected at a confidence level of 0.95. Thus, there is a trend indicating the relationship between the level of carcinogenicity and the value of the energy of affinity for the electron of the molecule. Apparently, the value of the energy of affinity can limit the carcinogenic activity of molecules - with a decrease in the energy of affinity, the bioactivity of molecules decreases at similar values of the factor Z . For molecules that have a range of affinity energies, their average value was used.

The pyrene molecule can be added to the series of compounds (31), which was formed by Pullman. Pyrene by factor $Z = 2.846$ arb. units is a potentially active carcinogen. However, it is possible that the activity of the pyrene is limited by the energy properties of the molecule. Indeed, the ionization potential of pyrene is relatively high, $I = 7.5$ eV, and the wavelength is shifted to the short-wavelength region $\lambda \approx 690$ nm. These values are closer to the areas that characterize carcinogenically inactive substances. Based on the electron capture experiment, the electron affinity energy of pyrene was determined $A = 0.50 \pm 0.03$ eV [44]. Thus, in

accordance with the values of the electronic factors, the pyrene molecule should have a relatively low carcinogenic activity.

Let us compose the following sequence of hydrocarbons with four to six condensed rings in order of decreasing their carcinogenic activity (this series is based on the sequence of hydrocarbons proposed by Pullman [35]):

Dibenz-3,4,8,9-pyrene ($Z = 2.895$ arb. units; $H = 0.95$ *bits*; $\gamma = 1.714$; $A = 1.00$ eV; +++) \approx dibenz-3,4,9,10-pyrene ($Z = 2.895$ arb. units; $H = 0.955$ *bits*; $\gamma = 1.714$; $A = 0.84$ eV; +++) \approx benz-3,4-pyrene ($Z = 2.875$ arb. units; $H = 0.953$ *bits*; $\gamma = 1.667$; $A = 0.77$ eV; +++) $>$ dibenz-1,2,3,4-pyrene ($Z = 2.895$ arb. units; $H = 0.95$ *bits*; $\gamma = 1.714$; $A = 0.87$ eV; +++) $>$ dibenz-1,2,5,6-anthracene ($Z = 2.833$ arb. units; $H = 0.964$ *bits*; $\gamma = 1.571$; $A = 0.68$ eV; ++) \geq dibenz-1,2,7,8-anthracene ($Z = 2.833$ arb. units; $H = 0.964$ *bits*; $\gamma = 1.571$; $A = 0.69$ eV; +) \approx dibenz-1,2,5,6-phenanthrene ($Z = 2.833$ arb. units; $H = 0.964$ *bits*; $\gamma = 1.571$; $A = 0.60$ eV; +) \approx dibenz-1,2,3,4-phenanthrene ($Z = 2.833$ arb. units; $H = 0.964$ *bits*; $\gamma = 1.571$; $A = 0.40$ eV; +) $>$ pyrene ($Z = 2.846$ arb. units; $H = 0.961$ *bits*; $\gamma = 1.6$; $A = 0.50$ eV; +/-) $>$ tetraphene ($Z = 2.80$ arb. units; $H = 0.971$ *bits*; $\gamma = 1.50$; $A = 0.70$ eV; +/-) \approx chrysene ($Z = 2.80$ arb. units; $H = 0.971$ *bits*; $\gamma = 1.50$; $A = 0.42$ eV; +/-) $>$ benz-1,2-pyrene ($Z = 2.875$ arb. units; $H = 0.955$ *bits*; $\gamma = 1.667$; $A = 0.07$ - 0.40 eV; -) \approx tetracene (naphthacene) ($Z = 2.80$ arb. units; $H = 0.971$ *bits*; $\gamma = 1.50$; $A = 0.80$ eV; -) \approx triphenylene ($Z = 2.80$ arb. units; $H = 0.971$ *bits*; $\gamma = 1.50$; $A = 0.28$ eV; -).

(33)

For the ranked series according to the value of carcinogenic activity, there is a trend for the value of factor Z . Abbe-Linnik test (32) will be as follows: $q = 0.519 < q_{0.05}^{\text{cr}}(N = 14) = 0.5908$. That is, an increase in molecular factor Z is accompanied by an increase in the carcinogenic activity of molecules. However, electronic sign A may be the limiting factor. For highly active molecules, the electron affinity is greatest. For example, for the first ten molecules (33), there is a significant trend for factor A : $q = 0.308 < q_{0.05}^{\text{cr}}(N = 10) = 0.5311$. A decrease in the electron affinity value is accompanied by a decrease in carcinogenic activity. At the same time, if we estimate the Spearman rank correlation coefficient for the ranked series by the value of the feature Z , then we obtain the following inequalities $\rho = 0.917 > \rho_{0.05}^{\text{cr}}(N = 14) = 0.4593$, $r = 0.924 > r_{0.05}^{\text{cr}}(N - 2) = 0.532$ [16]. That is, for this group of compounds, proposed by Pullman, there is a close relationship between the value of Z and the carcinogenicity of chemical compounds.

However, there are molecules with a sufficiently large value of factor A that do not have strong carcinogenic properties (tetracene, tetraphene). In this case, the factor Z is already the limiting factor, which is minimal for these molecules ($Z = 2.80$ arb. units). Such influence of the independent variables on the bioresponse the possibility of constructing multiple regression. A quantitative relationship between carcinogenicity levels and molecular factors can be obtained if the qualitative indicators for the Pullman series (10) are replaced with quantitative values, for example, points (++++ 4; +++ 3; ++ 2; + 1; +/- 0.5; - 0). In this case, we can write the following two-factor regression:

$$\text{Carc}(\text{points}) = b_0 + b_1 \cdot Z + b_2 \cdot A, \quad (34)$$

$N = 14$, $m = 2$; multiple correlation coefficient: $R = 0.905 > R_{0.05}^{\text{cr}}(f_1 = N - m - 1; f_2 = m) = 0.648$ [50]; $R^2 = 0.819$; $RMSE = 0.723$; $b_0 = -72.23 \pm 16.22$, $b_1 = 25.28 \pm 5.76$, $b_2 = 3.03 \pm 0.84$, $|t(b_0)| = 4.54 > t(b_1) = 4.40 > t(b_2) = 3.61 > t_{0.05}^{\text{cr}}(f = 11) = 2.201$; $F = 24.87 > F_{0.05}^{\text{cr}}(f_1 = 2; f_2 = 11) = 3.18$; standardized regression coefficients [51]: $b_1^* = 0.60$, $b_2^* = 0.50$; $\Sigma_0 = 5.746$.

Here the standard deviations of the populations Z and A are $S_Z = 0.037$, $S_A = 0.231$; Σ_0 is the sum of the squares of the regression residuals; m is the number of independent explanatory variables. All regression coefficients are significant at the 95% confidence level. Let's check the presence of collinearity between the explanatory variables. The linear correlation coefficient between the variables is $r_{1,2} = 0.44$. Since the regression residuals are normally distributed (Wilk-Shapiro normality test: $W = 0.945 > W_{0.05}^{\text{cr}}(N = 14) = 0.874$), we will use the Farrar-Glauber relation [52] to quantify significant collinearity:

$$\chi^2 = -(N - 1 - (2m + 5)/6) \cdot \ln \left(\det \begin{vmatrix} r_{1,1} & r_{1,2} \\ r_{2,1} & r_{2,2} \end{vmatrix} \right) =$$

$$2.47 < \chi_{0.05}^{2,\text{cr}}(f = 1) = 3.841. \quad (35)$$

It follows from inequality (35) that there is no significant collinearity between the explanatory variables. All populations: Carc, Z and A are homogeneous and have a distribution close to normal. In accordance with the values $b_1^* > b_2^*$ of the standardized regression coefficients (34), it can be noted that the explanatory variables Z and A make approximately the same contribution to the variability of the effective sign. Indeed, the approximate coefficient of determination makes it

possible to determine the share of explanatory variables in explaining the variability of the effective trait:

$$R_{\text{appr}}^2 = b_1^* \cdot r_{Z-\text{carc}} + b_2^* \cdot r_{A-\text{carc}} =$$

$$0.468 + 0.351 = 0.819. \quad (36)$$

Here $r_{Z-\text{carc}} = 0.78$, $r_{A-\text{carc}} = 0.71$ are the paired correlation coefficients of the explanatory variables (A, Z) with the effective indicator (Carc). The share of the explained variability of the effective trait from the variable Z is 46.8%, and the share of the variable A is 35.1%.

It can be tentatively assumed that the boundary (threshold) value of the affinity energy of hydrocarbons is determined by the energy interval $A^{\text{thr}} \approx 0.30\text{--}0.68$ eV, below which the manifestation of the carcinogenic activity of the molecule is substantially limited. This boundary value does not contradict the estimate (30). This area corresponds to the values $Z \leq 2.833$ arb. units. If the sample (33) is divided into two samples, one of which contains compounds with an activity of 2-4 points ($N_1 = 5$), the other contains drugs with an activity of 0-1 points ($N_2 = 9$), then the mean values of these samples in the range of values A^{thr} change by a jump. Quantitatively, the presence of a jump can be determined using the Cochran relation (30):

$$\chi^2 = (N_1 \cdot N_2 / N) (\text{Carc}_1^{\text{av}} - \text{Carc}_2^{\text{av}})^2 / \text{Carc}^{\text{av}} = 17.6 >$$

$$\chi_{0.05}^{2,\text{cr}}(f = 1) = 3.841. \quad (37)$$

Here Carc^{av} is the overall average. Since the inequality $\chi^2 > \chi_{0.05}^{2,\text{cr}}$ is fulfilled, then in the area of A^{thr} (and $Z \leq 2.833$ arb. units) there is a significant change in the carcinogenic activity of drugs.

That is, it can be assumed that at the initial stage of the action of a polycyclic hydrocarbon, the molecule is bound, then an electron is transferred, followed by the release of energy. The limitation of carcinogenic activity occurs both by the value of the factor Z and by the values of the energy of affinity and the first ionization potential. Unfortunately, the limited literature data on the values of the energies of the electron affinity of molecules does not allow statistically reliable determination of more accurate threshold values of this factor for polycyclic aromatic hydrocarbons.

For compounds with six condensed rings, Pullman indicated that dibenz-3,4,8,9-pyrene (++++), dibenz-3,4,9,10-pyrene (++++), and dibenz-1,2,3,4-pyrene (+++) are powerful carcinogens. For these hydrocarbons, the values of the factor attribute are really high: $Z = 2.895$ arb. units, and the electron affinity energy is in the region of high values: $A =$

1.1 eV, 0.80-0.84 eV, and 0.80-0.87 eV [44], respectively. That is, these electronic characteristics of molecules do not interfere with the manifestation of carcinogenic activity by drugs. According to limited data, carcinogens are anthanthrene ($Z = 2.941$ arb. units; $H = 0.937$ bits; $A = 1.0$ eV; $I = 7.10$ eV; $\lambda \approx 855$ nm), which has six condensed rings and perylene ($Z = 2.875$ arb. units; $H = 0.955$ bits; $A = 0.85$ eV; $I = 6.8$ eV; $\lambda = 920$ nm [36]) with five condensed rings. At the same time, for inactive dibenz-1,2,7,8-naphthacene, which also contains six condensed rings, factor Z has a relatively low value of 2.81 arb. units, although the affinity energy for this molecule is rather high, 0.9 eV [44]. In this case, factor Z is already the limiting factor of carcinogenic activity.

Thus, all the main provisions that were used by A. and B. Pullman as fundamental in their model can be interpreted within the framework of the model proposed here, which uses factors Z , H , γ , λ , I and A as explanatory variables.

It is possible to more definitely reveal the role of an exogenous molecule in electron transfer if the object or its active part with which the polycyclic molecule interacts is precisely known. As mentioned above, electron transfer is usually accompanied by endoenergetic or exoenergetic processes. These processes can affect the state of biological objects with which the polycyclic hydrocarbon interacts. Obviously, the participation of polycyclic aromatic molecules in reactions with electron transfer depends on the value of the electron affinity energy of the molecule. That is, the bioactivity of molecules having the same values of the molecular factors Z , H and γ , but different values of the electron affinity will be different. Since a carcinogen molecule can interact in the body with molecular structures, for example, in particular with RNA and DNA [53], the locally released energy during electron transfer can be directed to the destruction of these ordered molecular structures, thereby disrupting their normal functioning. The more exogenous molecules are involved in complexation, the stronger the total processes accompanying electron transfer.

It was established [12] with high accuracy that polycyclic aromatic compounds attach to DNA, and carcinogenic activity is proportional to the amount of hydrocarbons involved in complexation. Polycyclic aromatic hydrocarbons are usually planar molecules and they intercalate between DNA base pairs [54]. Complexes of DNA and flat molecules of acridine dyes are formed by a similar mechanism [55].

Molecules of nitrogenous bases - pyrimidine

derivatives (cytosine, thymine and uracil) and purine derivatives (adenine and guanine) are important components of nucleic acids. For all purine and pyrimidine bases, molecular factors differ markedly in magnitude from those of polycyclic aromatic hydrocarbons. The value of factor Z (or factor H) for purine and pyrimidine bases is significantly higher than for all known polycyclic aromatic hydrocarbons. The first ionization potential of nucleobases is significantly higher also than the ionization potential of carcinogenic aromatic hydrocarbons. For example, for guanine ($Z = 3.50$ arb. units) and adenine ($Z = 3.33$ arb. units), the ionization potentials are (8.3 ± 0.2) eV and (8.8 ± 0.2) eV [56], respectively. The electron affinity energy of these molecules differs significantly (not only in magnitude, but also in sign) from the affinity energy of polycyclic hydrocarbons and are equal for guanine $A = -0.96$ eV and for adenine $A = -0.77$ eV, respectively. For uracil and cytosine, the electron affinity energies, according to various data, are in the ranges of values $(+ 0.02, -1.12)$ eV and $(-0.24, -1.25)$ eV. It is possible that these differences in the electronic properties of nitrogenous bases and polycyclic hydrocarbons determine the selectivity of the binding of hydrocarbons to DNA regions. In accordance with inequalities (4), carcinogenic polycyclic hydrocarbons and nucleobases have the maximum difference in ionization potentials. For non-carcinogenic hydrocarbons, this difference is minimal.

As a result of complexation and subsequent energy processes caused by the transfer of an electron, the possibility of a violation of the transfer of genetic information is also not excluded. Apparently, the approximate condition $Z \geq Z^{\text{thr}}$ is a necessary but insufficient requirement for the carcinogenicity of classical polycyclic aromatic hydrocarbons. The molecule must also have appropriate energy characteristics: high affinity energy, low value of the first ionization potential, maximum wavelength of the absorption band due to charge transfer, which must be in the long-wavelength part of the spectrum. At the same time, only a high affinity energy does not guarantee that a molecule has carcinogenic properties if the factor Z (as well as H or γ) is close to its boundary value. For example, benz-1,2-naphthacene ($Z = 2.83$ arb. units, $A = 1.1$ eV), pentaphene ($Z = 2.83$ arb. units, $A = 1.2$ eV) and naphthacene ($Z = 2.80$ arb. units, $A = 0.88$ eV) have high values of electron affinity, but do not have confirmed carcinogenic activity.

Addition of an alkyl group to a polycyclic hydrocarbon can have a limiting effect on the

carcinogenicity of a substance [46]. First, this leads to a decrease in the magnitude of the molecular factors Z and H . Second, an increase in the length of the alkyl group is accompanied by a change in the hydrophobic properties of the substance, which prevents the manifestation of the biological activity of the chemical compound. According to A. and B. Pullman, steric obstacles can also be limiting factors.

In accordance with the model of A. and B. Pullman, chemical compounds LVIII and LIX (numbering suggested by Pullman [35]), which have a significant spatial size, should be carcinogenic. The same conclusion follows from the calculations of molecular factors, which are equal to $Z = 2.96$ and 3.00 arb. units, $\gamma = 1.88$ and 2.00 , respectively. However, it was not possible to find in the literature the results of an experimental test of the carcinogenicity of these substances. In addition, bioactivity depends on the method of application of the drug, tissue susceptibility, which is associated

with the species of the experimental animal, its age, sex, nutrition, harmonious state and the phase of mitotic activity of certain cells [1,2]. The activity can also be limited by the degree of binding of hydrocarbons to nucleic acids. The degree of binding depends on the conditions for the formation of the complex. These conditions depend on the dispersion of the hydrocarbon particles, the pH of the solution, and the ionic strength [19]. For example, the degree of binding decreases sharply with increasing ionic strength of the solution and increasing pH [31,37].

It can also be noted that the electronic factor Z , informational H and structural γ factors are highly informative. For example, Table 2 shows the experimental values of the energy ΔE of the most intense electronic transition [57] for several aromatic hydrocarbons for which the energies of electronic transitions are known.

Table 2
Electronic energies ΔE , δE and molecular factors of cyclic compounds.

Chemical compounds	ΔE , eV [58]	δE , cm ⁻¹ [45]	Z , arb.units	H , bits	γ
Benzene	4.70	43000	2.50	1.00	1.00
Naphthalene	3.94	28200	2.67	0.99	1.25
Anthracene	3.27	20700	2.75	0.98	1.40
Naphthacene	2.63	15500	2.80	0.97	1.50
Pentacen	2.16	11800	2.83	0.96	1.57

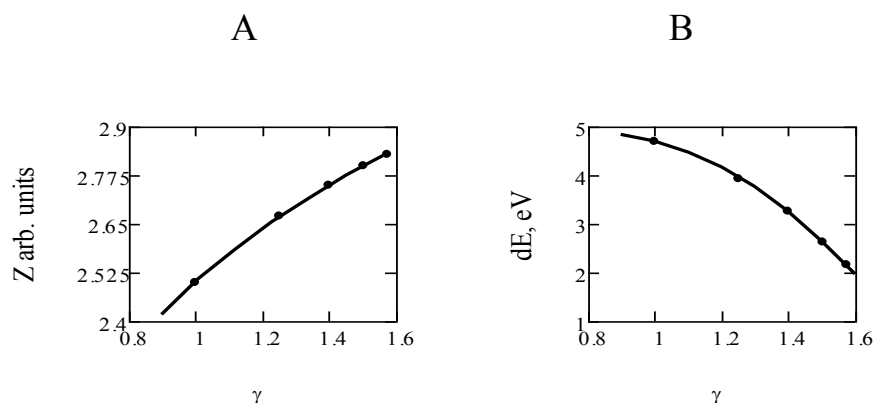


Figure 3. A. Relationship between factors Z and γ . Relationship statistics: $Z(\gamma) = B + A \exp(-C\gamma)$, $B = 3.23 \pm 0.03$, $A = -2.09 \pm 0.04$, $C = 1.05 \pm 0.06$, $R^2 = 0.999$, $RMSE = 0.0012$; Wilk-Shapiro test: $W(\gamma) = 0.938 > W(Z) = 0.901 > W_{0.05}^{cr}(N) = 0.767$; B. The relationship between the factors ΔE and γ . Relationship statistics: $\Delta E(\gamma) = B + A \exp(-C\gamma)$, $B = 5.77 \pm 0.13$, $A = -0.13 \pm 0.03$, $C = -2.13 \pm 0.14$, $R^2 = 0.999$, $RMSE = 0.0022$; Wilk-Shapiro test: $W(\Delta E) = 0.977 > W_{0.05}^{cr}(N) = 0.767$. For the ordinate axis, the notation $dE \equiv \Delta E$ is used.

From the data given in Table 2 and presented in Figures 3A and 3B, it is easy to see that the value of the electronic transition energy of cyclic

hydrocarbons is associated with the value of the structural factor γ of polyacene. Since the factor γ is closely related to the factor Z (Fig. 3A), the energy

of the electronic transition also obviously correlates with the value of the molecular factor Z . A small $RMSE$ and a very high coefficient of determination R^2 indicate that for this limited sample, the relationship approaches a functional relationship. Taking into account the physical meaning of molecular factors, it is possible to assume probable molecular processes in which polycyclic hydrocarbons are involved or with which the carcinogenic properties of active molecules are conjugative.

Let us also consider the relationship between the observed values of the delocalization energy (or resonance energy Er) of cyclic hydrocarbons and the value of their factor Z . The delocalization energy is usually taken as the difference between the heat of combustion of a hydrocarbon and the heat of combustion, which is calculated on the basis of the assumption of the additivity of the binding energies of molecules [48].

In this case, the Z factor turns out to be informative for polycyclic hydrocarbons,

demonstrating (Fig. 4A) the linear relationship of the Z value with the resonance energy value:

$Er(Z) = a + b \cdot Z$, $N = 10$, $R = 0.96 \pm 0.10$; $R^* = 0.965 > R_{0.05}^{cr}(N-2) = 0.632$, $a = -624.18 \pm 77.32$, $b = 261.09 \pm 28.05$, $t(b) = 9.31 > |t(a)| = 8.03 > t_{0.05}^{cr}(N-2) = 2.306$; $RMSE = 8.76$; significance criterion (2) for the correlation coefficient of the Fisher z -transform (taking into account Hotelling's corrections [16]): $u_H = 2.014 > u_{0.05}(N) = z_{0.975} \cdot (N-1)^{-0.5} = 0.65$; $F = 86.7 > F_{0.05}^{cr}(f_1 = 1; f_2 = 8) = 5.32$; straightforwardness: $K = (N \cdot (1 - R^2))^{0.5} = 0.89 < K^{thr} = 3.00$ [15]. (38)

For small sample sizes ($N < 30$), the correlation coefficient must be adjusted in accordance with the ratio (3). Figure 4B demonstrates the relationship between the observed excitation energy δE (in units of cm^{-1}) [45] for transitions from the highest occupied energy level to the lowest free level with the factor Z in polyacene molecules.

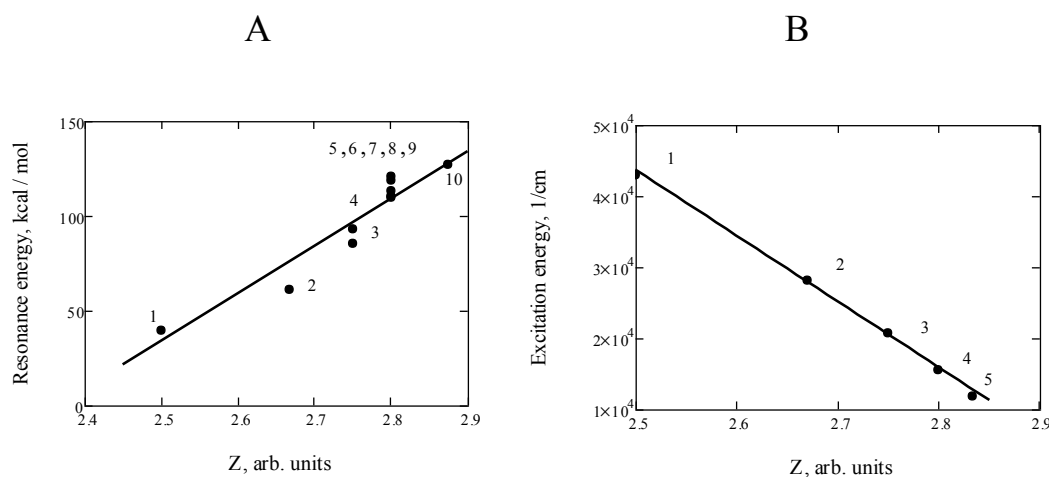


Figure 4. A. Interrelation of the resonance energy of Er [48] cyclic hydrocarbons with the molecular factor-sign Z . 1 – benzene, 2 – naphthalene, 3 – anthracene, 4 – phenanthrene, 5 – benz-3,4-anthracene, 6 – tetracene, 7 – pyrene, 8 – benz-1,2-anthracene, 9 – chrysene, 10 – perylene. B. Interrelation of energy of transition from the highest filled energy level to the lowest free level [45] with the factor Z for cyclic hydrocarbons. 1 – benzene, 2 – naphthalene, 3 – anthracene, 4 – naphthacene, 5 – pentacene.

The statistics of the relationship of factors is as follows:

$$\delta E(Z) = a + b \cdot Z, N = 5, R = -0.999 \pm 0.030, |R^*| =$$

$$0.9995 > R_{0.05}^{cr}(N-2) = 0.8783, a = 276941.6 \pm$$

$$7608.1, b = -93396.7 \pm 2804.8, t(a) = 36.40 > |t(b)| = 33.30 > t_{0.05}^{cr}(N-2) = 3.182; F = 1108.8 > F_{0.05}^{cr}(f_1 = 1; f_2 = 3) = 10.13; \text{Wilk-Shapiro normality test: } W = 0.901 > W_{0.05}^{cr}(N) = 0.762; \text{straightforwardness: } K = (N \cdot (1 - R^2))^{0.5} = 0.10 < K^{thr} = 3.00. (39)$$

Thus, the electronic characteristic δE is related to the Z factor, and the relationship is practically functional. There is a similar close relationship for

the explanatory factor

The high information and substantive content of factors Z and H is also indicated by their parallelism with donor activity not only for polycyclic hydrocarbons, but also for other classes of chemical compounds. For example, an experimental results of the quadrupole splitting of the gamma-resonance line of tin in the Mössbauer spectrum of dibutyl tin chloride [58] with respect to the organometallic compound $(C_4H_9)_2SnCl$ for seven solvents confirms this relationship.

It is important to emphasize that the model proposed here is not only due to the properties of the electrons of the outer shell of the atoms of the entire molecule, but also establishes the conjugation between the various physical properties of the molecules as a whole. The importance of taking into account the properties of molecules as a whole when assessing the carcinogenicity of chemical compounds was understood by A. and B. Pullman, who wrote "... that the significant reactivity of a number of highly carcinogenic compounds in some specific reactions cannot be considered a sufficient basis for creating a theory linking carcinogenic activity with chemical reactivity" [35].

In our opinion, it is fundamentally important for the compared models that, in contrast to the Pullman's model, the model proposed here uses molecular factors (Z , H , γ), which, firstly, are obtained from various independent principles, and secondly, these factors characterize the molecule as a whole.

The stated statistical approach to the selection of chemical compounds using molecular factors Z , H , γ establishes classification rules that allow us to identify the similarities and differences between the variable molecular structures of drugs and the carcinogenic effect of drugs. The statistical approach proposed here can be practically useful in preliminary prediction of the carcinogenicity of new polycyclic compounds. It is important to emphasize that rather simple calculations of the molecular factors Z , H and γ require knowledge of only the chemical gross formula of the molecule. The ability of factors to separate potentially carcinogenic compounds from non-carcinogenic drugs, apparently, is not accidental and is associated with the actually existing electrostatic molecular potential, the value of which varies from molecule to molecule.

4 Conclusion

The change in the carcinogenic properties of polycyclic hydrocarbons with an increase in the factor Z (or factor H) does not contradict the known

concepts of the mechanisms of chemical carcinogenesis. The known data [19], as well as quantum-chemical calculations [59], allow us to conclude that, at least in a series of structurally similar chemical compounds, the carcinogenicity property is associated with their participation in the donor-acceptor interaction. However, this relationship can be leveled by a change in the relative position of free and occupied molecular orbitals of molecules. The molecular potential, in turn, through the electrostatic field created by it, affects the processes that regulate the vital activity of a biological object, thereby determining the bioactivity of chemical compounds.

Thus, two different approaches to assessing the carcinogenic activity of polycyclic hydrocarbons, based on different principles, generally do not contradict each other. At the same time, the readily available method for calculating the factors Z , H and γ does not require complex and laborious quantum-chemical calculations of the electronic structure of molecules, the knowledge of which is necessary for a quantitative assessment of the properties of the K and L regions of a polycyclic hydrocarbon. Molecular factors, on the one hand, can be used for rapid assessment of the carcinogenicity of chemical compounds, and on the other hand, the physicochemical meaning that is embedded in these factors allows one to presumably indicate the physicochemical processes in which bioactive molecules can participate.

References:

- [1] R.Schoental, *Supplement to the book Bedger G.M. Chemical bases of carcinogenic activity*. M. Medicine. 1964. 124 p. (Russian edition).
- [2] R.Schoental, *Nature (London)*, Vol. 192, 1961, p. 670.
- [3] O.Chalvet, R.Daudel, *Compt. Rend.*, 242 D, 1956, pp.416-418.
- [4] N.Muller, L.C.Pickett, R.C.Mulleken, *J. Am. Chem. Soc.*, Vol.76, 1954, pp.4770-4778.
- [5] A.Pullman, B.Pullman, In: *Advances in Cancer Research*. Eds. by Greenstein J., Haddow A. Vol. 3, 1955. Acad. Press. Inc. Publisher. New-York.
- [6] J.Koutecky, R.Zahradnik, *Cancer Research*, Vol.21(4), 1961, pp.457-462.
- [7] S-S.Sung, *C.R. Acad. Sci. Paris*, 274 D, No.10, 1972, pp.1597-1600.
- [8] Coulson C.A. *Adv. Cancer. Res.*, Vol. 1, 1953, p.1.
- [9] B.Pullman, *La Biochimie Electronique*. Press Universitaires de France. 1963.
- [10] V.K.Mukhomorov, *SCIREA Journal of*

Biology, Vol.3, 2018, pp.1-14.

- [11] V.K.Mukhomorov, *Biomedical Statistics and Information*, Vol.1, No.3, 2016, pp.24-34.
- [12] P.Brookes, P.D.Lawley, *Nature*, Vol.202, No. 4934, 1964, 781.
- [13] J.Caillet, J.M.Thomas, *Molecular Associations in Biology*. Ed. B.Pullman. Academic Press, 1968.
- [14] R.Rein, In: *Intermolecular Interaction. From Diatomics to Biopolymers*. Ed. B. Pullman, A Wiley-Interscience Publication, Chichester-New York-Brisbane-Toronto, 1981.
- [15] G.N.Zaitsev, *Mathematics in Experimental Botany*, Moscow, Science, 1990. (in Russian).
- [16] L.Sachs, *Statistische Auswertungsmethoden*. Springer-Verlag, Berlin, Heidelberg, New-York, 1972.
- [17] A.I. Kobzar, *Applied Mathematical Statistics. For Engineers and Scientists*, Moscow, Fizmatlit, 2016. (in Russian).
- [18] M.Stamatiadou, T.J.Swissler, R.Rein, *Biopolymers*, Vol. 11(6), 1972, pp.1217-1234.
- [19] M.M. Vilenchik, *Regularities of the molecular-genetic action of chemical carcinogens*, Moscow, Science, 1977. (in Russian).
- [20] E.D.Bergman, In: *Molecular Association in Biology*, New York, Acad. Press, 1968.
- [21] A.R.Peacocke, *Biopolymers*, Vol.6, 1968,1225.
- [22] E.Boyland, B.Green, *Biochem. J.*, Vol.92, 1969, 4.
- [23] A.M.Craig, I.Isenberg, *Proc. Nat. Acad. Sci. USA*, Vol.67, 1970, 1337.
- [24] B.Green, *Biochem. J.*, Vol.112, 1969, 183.
- [25] J.J.Wang, E.E.Rasmussen, T.T.Crocker, *Biochem. and Biophys. Res. Commun.*, Vol.49, 1972, 1142.
- [26] S.Lesco, H.D.Hoffman, P.O.Nso, V.M.Maher, *Progress in Molecular and Subcellular Biology*, Vol.2, New York, Acad. Press, 1971.
- [27] S.Lesco, P.O.Nso, R.S.Man, *Biochemistry*. Vol.7, 1968, 434.
- [28] Y.Pasca, F.Pachon, A.M.Michelson, *Biochemica*, Vol.53, 1971, 365.
- [29] J.Clifford, K.Ress, M.E.Stevens, *Biochem J.*, Vol.103, 1967, 258.
- [30] J.Schabart, *J. South Afric. Inst.*, Vol.22, 1968, 80.
- [31] E.Boyland, B. Green, *Brit. J. Cancer.*, Vol.16, 1964, 507.
- [32] M.Green, *Anual Rev. Biochem.*, Vol.39, 1970. 701.
- [33] N.L.Johnson, F.C.Leone, *Statistics and Experimental Design in Engineering and the Physical Sciences*, Vol.1, John Wiley&Sons, New York-London-Sydney-Toronto, 1977.
- [34] E.Förster, B.Rönz, *Metohden der Korrelations- und Regressionsanalyse*, Verlag Die Wirtschaft, Berlin, 1979.
- [35] A. Pullman, B. Pullman, *Collected papers, Advances in cancer research*. Vol..3, Moscow, IL, 1957. (Russian edition).
- [36] *Handbook. Carcinogenic substances*. Ed. V.S.Turusov. Moscow, Medicine, 1987. (IARC Monographs on the Evaluation of the Cancerogenic Risk of Chemical Compounds. Intern. Agency for Research on Cancer). (in Russian).
- [37] B.Green, *Europ. J. Biochem.*, Vol.14, 1970, 567.
- [38] L.S.Lerman, *Proc. 5th Nat. Cancer Conf.*, p.39, 1964.
- [39] K.N.Prosad, *Nature. New Biol*, Vol.236, 1972, 43.
- [40] P.N.Dyachkov, *Quantum-chemical calculations in the study of the mechanism of action and toxicity of substances. Results of Science and Technology. Ser. Toxicology*, Vol. 16, Moscow, VINITI, 1990. (in Russian).
- [41] G.M.Bedger, *The Chemical Basis of Carcinogenic Activity. Supplements to the Russian edition*, Moscow, Medicine, 1964. (Russian edition).
- [42] N.I.Vylegzhanin, Experience in studying the blastomogenic properties of 3,4,6,7-dibenzpyrene. Collection of articles, *Problems of Oncology*, Vol. 21, No. 5, 1952, pp.135-140. (in Russian).
- [43] G.M.Badger, *J. Chem. Soc.*, 1809, 1960.
- [44] L.V.Gurevich, G.V.Karachevtsev, V.N. Kondrat'ev, Yu.A. Lebedev, V.A. Medvedev, V.K. Potapov, Yu. S. Khodeev, *The Energies of Breaking Chemical Bonds. Ionization Potentials and Electron Affinity*, Moscow, Science, 1974. (in Russian).
- [45] Keniti Higasi, Hiroaki Baba, Alan Rembaum, *Quantum Organic Chemistry*, Interscience Publishers John Wiley & Sons, New York, London, Sydney, 1965.
- [46] L.F.Fieser, S.T.Putman, *J. Am. Chem. Soc.*, Vol.69, 1947, 1041.
- [47] J.Ladik, *Quantenbiochemie für Chemiker und Biolgen*, Akadimiai Kiado Budapest, 1970.
- [48] J.N.Murrell, S.F.A.Kettle, J.M.Tedder, *Valence Theory*, John Wiley & Sons, New York, London, Sydney, 1965.
- [49] A.A.Maevsky, M.M.Vilenchik, B.I.Sukhorukov, *Biophysics*, Vol.18, 1973, 371 (in Russian).
- [50] Jiří Likeš, Josef Laga, *Základní Statistické Tabulky*, Praha, SNTL-Nakladatelství Technické Literatury, 1978.

- [51] E.A.Dmitriev, *Mathematical Statistics in Soil Science*, Moscow, Ed. Moscow State University, 1995. (in Russian).
- [52] D.E.Farrar, R.R.Glauber, *The Review of Economics and Statistics*, Vol.49, No.1, 2018, pp.92-107.
- [53] C.Heidelberger, G.R.Devanport, *Acta Unio Contra Concrum.*, Vol.17, 1961, 55.
- [54] A.M.Liquori, B.Lerma, F.Ascoli, C.Botr, *J. Mol. Biol.*, Vol.5, 1962, 511.
- [55] P.Alexander, *Nature*, Vol.235, 1971, 137.
- [56] I.I.Shafran'osh, Yu.Yu.Svida, M.I.Sukhovia, M.I.Shafran'osh, V.A.Mineev, G.V. Baryshnikov, *Journal of Technical Physics*, Vol. 85, No.10, 2015, 16 (in Russian).
- [57] T.E.Peacock, *Electronic Properties of Aromatic and Heterocyclic Molecules*, Academic Press, New York, London, 1965.
- [58] V.I.Goldanskii, *Angew. Chem.*, Vol. 79, No.19, 1967, 844.
- [59] B.Pullman, A.Pullman, *Quantum Biochemistry*, Interscience Publisher John Wiley & Sons, New York, London, 1962.
- [60] Ya. L. Kostyukovsky, D. B. Melamed, Carcinogenic N-Nitrosamines. Education. Properties. Analysis. *Advances in Chemistry*, Vol. LVII, No. 4, 1988, pp.625-655. (in Russian).
- [61] V.J.Veljković, D.J.Lalović, *Cancer Biochem. Biophys.*, Vol. 1, 1976, 295.
- [62] A.D'Albis, M.P.Wickens, W.P.Gratzer, *Biopolymers*, Vol. 14, 1975, 1425.
- [63] V.K.Mukhomorov, *International Journal of Chemical and Biomedical Science*, Vol. 3, No.4, 2017, 39.

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