Table 17. *Additive components of the pair interaction tnergy of tetramethyl-uric acid with polycyclic hydrocarbons.*

N	Chemical compound	Brutto formula	Contributions to the energy of pair interaction, kcal/mol [32]					$Z^{*)}$	$H^{*)}$, *bits*
			E_{elect}	E_{pol}	E_{disp}	E_{rep}	E_{total}		
1	Dibenz-1,2,3,4-pyrene	$C_{24}H_{14}$	-1.93	-1.11	-10.37	4.95	-8.46	2.90	0.95
2	Anthanthren	$C_{22}H_{12}$	-1.34	-0.93	-8.72	4.79	-6.20	2.94	0.94
3	Perylene	$C_{20}H_{12}$	-1.35	-0.86	-7.80	4.25	-5.76	2.88	0.96
4	Benz-1,2-pyrene	$C_{20}H_{12}$	-1.77	-0.96	-9.23	4.51	-7.45	2.88	0.96
5	Benz-3,4-pyrene	$C_{20}H_{12}$	-1.38	-0.86	-8.64	4.48	-6.40	2.88	0.96
6	Pyrene	$C_{16}H_{10}$	-1.46	-0.78	-7.76	4.12	-5.88	2.85	0.96
7	Dibenz-1,2,5,6-anthracene	$C_{22}H_{14}$	-1.09	-0.79	-8.23	3.88	-6.23	2.83	0.96
8	Chrysene	$C_{18}H_{12}$	-1.18	-0.80	-7.92	4.15	-5.75	2.80	0.97
9	Benz-1,2- anthracene	$C_{18}H_{12}$	-1.18	-0.80	-7.92	4.15	-5.75	2.80	0.97
10	Phenantrene	$C_{14}H_{10}$	-1.57	-0.65	-7.13	3.77	-5.58	2.75	0.98
11	Anthracene	$C_{14}H_{10}$	-1.27	-0.79	-6.84	3.67	-5.22	2.75	0.98

*) The elements of the sets Z and H satisfy the normal distribution. Statistics of Wilk-Shapiro: $W_Z = 0.94 > W_H = 0.92 > W_{0.95}^{(cr)}(f = 11) = 0.85$.

Statistical analysis has shown that there is a reliable relationship between the descriptor Z with the contribution to the energy of short-range repulsion, the polarization contribution, and the dispersion interaction. We have obtained the following correlation equation for the relationship of the descriptor Z with the energy E_{rep}:

$$E_{rep}(Z) = a_0 + a_1 Z, \quad N = 11, \quad R = 0.88,$$
$$a_0 = -12.28 \pm 3{,}00, \quad a_1 = 5.82 \pm 1.06,$$
$$F = 30.3 > F_{0.05}^{(cr)}(f_1 = 1; f_2 = 9) = 5.1;$$
$$W_{Erep} = 0.96 > W_{0.95}^{(cr)}(f = 11) = 0.85. \qquad (59)$$

Here f is the number of freedom degrees. The statistical significance of the correlation coefficient is confirmed by the inequality: $t = 0.5 \cdot \ln[(1+|R|)/(1-|R|)] \cdot (N-3)^{1/2} = 3.89 > t_{0.95}^{(cr)}(f = N-2) = 2.26$. The relationship between the contributions E_{pol} and E_{disp} with the molecular descriptor Z has the following statistics:

$$E_{pol}(Z) = a_0 + a_1 Z, \quad N = 11, \quad R = -0.77,$$
$$a_0 = 3.48 \pm 1{,}21, \quad a_1 = -1.52 \pm 0.43,$$
$$F = 12.7 > F_{0.05}^{(cr)}(f_1 = 1; f_2 = 9) = 5.1; \quad W_{Epol} = 0.93 > W_{0.95}^{(cr)}(f = 11) = 0.85, \qquad (60)$$

the statistical significance of the correlation coefficient R: $t = 2.85 > t_{0.95}^{(cr)}(f = 9) = 2.26$.

$$E_{disp}(Z) = a_0 + a_1 Z, \quad N = 11, \quad R = -0.74, \quad a_0 = 25.3 \pm 10{,}3, \quad a_1 = -11.8 \pm 3.62,$$
$$F = 10.7 > F_{0.05}^{(cr)}(f_1 = 1; f_2 = 9) = 5.1; \quad W_{Edisp} = 0.95 > W_{0.95}^{(cr)}(f = 11) = 0.85, \qquad (61)$$

the statistical significance of the correlation coefficient R: $t = 2.66 > t_{0.95}^{(cr)}(f = 9) = 2.26$.

Apparently, the molecular descriptor Z is allowed intermolecular interaction to be identified and described in accordance with the specification short-range components of intermolecular interactions that determine the formation of complexes of organic substances with molecular systems that are simulating biosystems. It is important to emphasize that the relationship exists for the dispersion interaction, which gives the dominant contribution to the total interaction energy. As the analysis showed, a linear relationship is absent for long-range

electrostatic interactions. However, it is necessary to note the following important fact. The detection of the interrelations (59) - (61) is not unexpected. Previously, we found similar relationships for other classes of chemical compounds. For example, there is a close interrelation between Z descriptor (and also H descriptor) and the absolute value of the dispersion interaction energy for halogenated compounds [9]:

$$E_{\text{disp}}(Z) = a_0 + a_1 Z, \quad N = 12, \quad R = 0.96,$$
$$a_0 = -1.10 \pm 0{,}95, \quad a_1 = 2.52 \pm 0.24,$$
$$F = 108.50 >> F_{0.05}^{(\text{cr})}(f_1 = 1; f_2 = 10) = 5.00;$$
$$W_{\text{Edisp}} = 0.95 > W_{0.95}^{(\text{cr})}(f = 12) = 0.86, \qquad (62)$$

the statistical significance of the correlation coefficient: $t = 5.83 > t_{0.95}^{(\text{cr})}(f = 10) = 2.23$.

Statistical analysis showed a close relationship between the molecular descriptors Z and H. However, polycyclic hydrocarbons have peculiarity in comparison with other chemical compounds of different classes (see Sect. 4). For polycyclic aromatic hydrocarbons, the relationship has a statistically significant opposite direction:

$$H(Z) = a_0 + a_1 \cdot Z, \quad N = 11, \quad R = -0.96,$$
$$a_0 = 1.49 \pm 0.05, \quad a_1 = -0.186 \pm 0.02,$$
$$RMSE = 0.0035, \quad t(a_0) = 29.6 > |t(a_1)| = 10.48 > t_{0.05}^{(\text{cr})}(f = 9) = 2.26,$$
$$F = 109.98 > F_{0.05}^{(\text{cr})}(f_1 = 1, f = 9) = 5.1, \qquad (63)$$

the statistical significance of the correlation coefficient: $t = 5.83 > t_{0.95}^{(\text{cr})}(f = 10) = 2.23$.

Statistical analysis confirms the opposite trend in the interrelationship between the descriptors, even if a wider range of polycyclic hydrocarbons are used (N = 41 [29], R = -0.996, $RMSE$ = 0.0012). Set of elements Z and H satisfy the normal distribution. An increase the sample size leads to the appearance of a weak nonlinearity (Fig. 8A). In this case, the $RMSE$ decreases to the value 0.0005. That is, the interrelationship is approaching to a functional dependence. The nonlinearity of the interrelationship occurs for most classes of chemical compounds [9].

It follows from the statistics (57) - (61) that the molecular descriptors Z and H are statistically significantly informative in order to characterize the pair intermolecular interaction. We emphasize that the descriptors have been derived of various principles. We will use these molecular descriptors to evaluate the carcinogenicity of polycyclic hydrocarbons.

The most active carcinogens are molecules that have the greatest value of the molecular descriptor Z. For example, dibenz-3,4,9,10-pyrene and dibenz-3,4,8,9-pyrene, for which descriptor Z is equal to 2.895. Taking into account the Table 17 and the relationships (57) - (60), it can be assumed that molecules with a high value of Z are preferred in paired dispersion interactions. Intermolecular interactions can only be a prerequisite for the appearance of complexes. Strong covalent chemical interactions might arise at the final stage of complex formation. This situation was indeed observed in the complexation of dibenz-1,2,5,6-anthracene with the cellular receptor [29].

Here we will analyze the experimental data that were used by Pullman A. and Pullman B. [29, 33] for interpretation of carcinogenic activity of polycyclic aromatic hydrocarbons. Pullmans [33] highlight the following group of polycyclic hydrocarbons, which contains from four to five condensed benzene rings: A very strong carcinogen – benz-3,4-pyrene (++++); moderately carcinogenic – dibenz-1,2,5,6-anthracene (++) and weak carcinogens – dibenz-1,2,3,4-phenanthrene (+), dibenz-1,2,7,8-anthracene (+); inactive – triphenylene (–) [5].

The sequence of values molecular descriptor Z will be as follows: 2.875; for the three penultimate compounds – 2.830; for the last triphenylene – 2.800. There is a parallelism between the carcinogenic activity of polycyclic aromatic hydrocarbons and the values of their molecular descriptor Z.

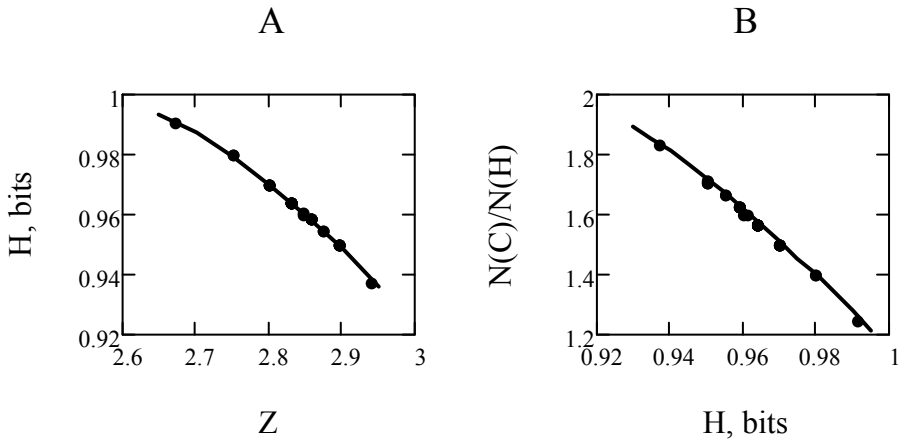

Fig.8. The polycyclic aromatic hydrocarbons. (A) *Interrelation of molecular descriptors H and Z.* (B) *Interrelation of molecular descriptors of* $\gamma = N(\mathrm{C})/N(\mathrm{H})$ *and H.*

Pullman proposed [29] distinguish three groups of carcinogenic substances according to their relative activity. The first group (inactive or weakly active) is benz-3,4-phenanthrene (Z = 2.80, H = 0.97*bits*), dibenz-1,2,5,6-phenanthrene (Z = 2.83, H = 0.96*bits*), dibenz-1,2,3,4-phenanthrene (Z = 2.83, H = 0.96*bits*); the second group (active) is dibenz-1,2,5,6-anthracene (Z = 2.83, H = 0.96*bits*), dibenz-1,2,7,8-anthracene (Z = 2.83, H = 0.96*bits*); the third group (the most active carcinogens) is benz-3,4-pyrene (Z = 2.875, H = 0.955*bits*), dibenz-1,2,3,4-pyrene (Z = 2.895, H = 0.95*bits*), dibenz-3,4,8,9-pyrene Z = 2.895, H = 0.95*bits*), dibenz-3,4,9,10-pyrene (Z = 2.895, H = 0.95*bits*). Thus, in this case we can also make a comparative assessment of the carcinogenicity for each cyclic hydrocarbon group using the molecular descriptor of Z (as well as the descriptor H; Fig. 8A). According to the level of carcinogenicity, three groups of compounds maintain their consistency. The greater the value of the descriptor Z, the more likely the intermolecular interaction is stronger.

Thus, the use of molecular descriptors Z and H does not contradict the criteria of carcinogenicity suggested by Pullmans, which were called the K and L regions of the molecules [33]. For example, dibenz-1,2,3,4-pyrene, dibenz-3,4,8,9-pyrene and dibenz-3,4,9,10-pyrene do not have the L region, but they have a suitable index of K region and therefore they are carcinogens. The molecular descriptors Z = 2.895 and H = 0.95*bits* also indicate the carcinogenicity of these compounds. Descriptors not just indicate, but the descriptors correctly determine the sequence in the relative activity of this series of compounds. Pullman also pointed out [29] that the criteria have exceptions for polycyclic hydrocarbons. First, benz-3,4-phenanthrene, dibenz-1,2,5,6-phenanthrene and dibenzo-1,2,3,4-phenanthren do not subject to the criteria of Pullmans [33]. Perhaps this is due to the inaccuracy of the established threshold values for the criteria of the L and K regions. As shown above, the use as a quantitative measure of the molecular descriptor Z = 2.80-2.83 does not deny the weak carcinogenicity of these compounds. This is also confirmed by the test for the promotive activity of mice for benz-3,4-phenanthrene. The test was positive [5]. Secondly, according to Pullmans, anthanthrene should have carcinogenic activity. The value of descriptor Z = 2.94 to point to this. According to modern data [5], anthanthrene is indeed of carcinogenic activity. The strong carcinogen the dibenz-3,4,6,7-pyrene does not have K region. At the same time, the molecular descriptor for this polycyclic hydrocarbon is very high (Z = 2.895) and hence this substance must be a carcinogen. This result does not contradict observations.

Thus, two different models based on different principles do not lead to contradictory results. At the same time, it is not necessary to perform complex and labor-intensive quantum mechanical calculations to determine the descriptors Z and H. It is enough to know only the gross formula of a polycyclic hydrocarbon. To determine the quantitative characteristics of the K and L regions, knowledge of the various

quantum chemical parameters of the molecules is required.

We note further that the most studied addition reactions occurring in the *L* region of the molecule are (according to Pullmans [33]): *a*) fixation of maleic anhydride and *b*) photochemical oxidation. The descriptor *Z* (or the descriptor *H*) enable tracking the facility of fixing maleic anhydride. Pullmans [33] indicated that the facility of fixation is consistently decreasing in the following series of compounds:

Anthracene ($Z = 2.75$) > benz-1,2-anthracene ($Z = 2.80$) > dibenz-1,2,5,6-anthracene ($Z = 2.83$) > benz-3,4-pyrene ($Z = 2.875$). (64)

Obviously, this sequence of fixations correlates with the value of descriptor *Z*. This gradation of the ease of fixation is confirmed by experience.

Photochemical oxidation very easily occurs in non-carcinogenic anthracene molecules ($Z = 2.75$) and naphthacene ($Z = 2.80$). Photooxidation becomes the more difficult with an increase of the number of side rings. Dibenz-1,2,5,6-anthracene molecule ($Z = 2.83$) weakly react, and benz-3,4-pyrene molecule ($Z = 2.875$) is completely inactive in the photooxidation reaction. Methyl substitution of the hydrocarbon increases the reactivity of the chemical compound. For example, 9,10-dimethylbenzanthracene ($Z = 2.67$, $H = 0.99$*bits*) is much more active than the initial hydrocarbon benzanthracene ($Z = 2.80$, $H = 0.97$*bits*), in reactions with maleic anhydride and reaction of photooxidation. These experimental facts also correlate with the variation of the molecular descriptor *Z* (or *H*). Indeed, benz-1,2-pyrene molecule ($Z = 2.875$) possesses a high carcinogenic activity [34], as well as the following polycyclic hydrocarbons (in order of increasing activity): dibenz-1,2,3,4-pyrene ($Z = 2.895$), dibenz-3,4,8,9-pyrene ($Z = 2.895$), dibenz-3,4,9,10-pyrene ($Z = 2.895$), and also ovalene ($Z = 3.087$). The ovalene has pronounced carcinogenic, mutagenic and teratogenic properties. As demonstrated in the article [35], the dibenz-3,4,6,7-pyrene is very active (no less active than the benz-3,4-pyrene or the dibenz-3,4,8,9-pyrene). The dibenz-3,4,6,7-pyrene has the descriptor $Z = 2.895$. At the same time it is important to note that dibenz-3,4,6,7-pyrene does not have a sufficiently active *K* region.

The addition reaction of osmium tetraoxide was investigated in detail to study the reactivity of the *K* region. The following sequence of polycyclic compounds has been drawn up [29] in order of increasing their reactivity with respect to this agent:

Phenanthrene ($Z = 2.75$; $I = 7.75$ eV; (–)) < benz-1,2-anthracene ($Z = 2.80$; $I = 7.50$ ev; (+/–)) < dibenz-1,2,5,6-anthracene ($Z = 2.83$; $I = 7.09$-7.80 eV; (++) < benz-3,4-pyren ($Z = 2.875$; $I = 7.19$ eV; (++++)). (65)

It is not difficult to see that in this case there is also a parallelism of the reactivity with the magnitude of the descriptor *Z*. The relative carcinogenic activity of molecules increases in the same sequence. Here we give the values of the first ionization potentials (*I*) of the molecules [36]. The increase in descriptor *Z* is attended by a decrease in the ionization potential of the molecule for a given series (65) of polycyclic hydrocarbons. There is the interrelation (65) between the magnitude of descriptor *Z* and the magnitude of the first ionization potential of polycyclic aromatic hydrocarbons. It can be assumed that the lower the ionization energy of the molecule, the better the donor properties of the molecule. The increase in the descriptor *Z* often coincides with the dynamics of the increase in the donor ability of molecules. Obviously, for the sequence of observations (65), the descriptor *Z* and the associated signs of *H* and *I* are reliably informative. The variation of a number of chemical compounds (65) indicates the existence of a relationship between the reactivity of molecules and the magnitude of molecular features. The descriptor value $Z = 2.80$-2.83 we can approximately taken as the boundary value in accordance with the sequence (65). This value approximately separates carcinogens from inactive polycyclic hydrocarbons. In this connection, we note that two dibenzphenanthrenes XIII and XIV (the numbering of Pullmans [33]) does not contain the reaction *K* region. Consequently, they cannot be attributed to the carcinogens. However, the molecular descriptor ($Z = 2.83$) indicates that these chemical compounds should have a weak carcinogenic activity. This is confirmed by observations [5]. Molecular descriptors are also useful for quantitatively assessing the reactivity of the *L* region of molecules. The order of decreasing reactivity towards $Pb(OAc)_4$ is :

Anthracene (Z = 2.75, H = 0.98*bits*; (–)) > benz-1,2-anthracene (Z = 2.80, H = 0.97*bits* (–/+)) > dibenz-1,2,5,6-anthracene (Z = 2.83, H = 0.96*bits* (+)) = 0. (66)

Molecular descriptors indicate their parallelism with the reactivity of the L region and in this case.

Another molecular descriptor can be proposed for the purpose of rapid assessment of the carcinogenic activity of polycyclic aromatic hydrocarbons. This descriptor is defined by the ratio of the number of carbon atoms to the number of hydrogen atoms: $\gamma = N(\mathrm{C})/N(\mathrm{H})$. As shown by statistical analysis, the descriptor γ is closely related to the molecular descriptors Z and H (Fig. 8B). There is a statistically significant close relationship between the descriptors for the forty-one polycyclic aromatic hydrocarbons represented by Pullmans [33]. For example, the relationship between the information function H and the descriptor γ can be approximated by a linear form:

$$\gamma(H) = a_0 + a_1 \cdot H, \quad R = -0.997, \quad N = 41,$$
$$RMSE = 0.0079, \quad a_0 = 11.86 \pm 0.13,$$
$$a_1 = -10.67 \pm 0.14, \quad |t(a_1)| = 78.6 >> t_{0.05}^{(\mathrm{cr})}(f = 34) = 2.03. \quad (67)$$

The increase in the sample size leads to the appearance of a statistically significant weak nonlinearity for the relationship between the signs of γ and H (Fig. 8B). A similar close relationship exists between the descriptors Z and γ. It is important to note that the molecular descriptors that were obtained on the basis of different baseline principles are closely interrelated. Their application leads to identical results for polycyclic aromatic hydrocarbons.

The above experimental facts can also be interpreted using the molecular descriptor γ, which is in no way connected with the K and L regions of the polycyclic molecule. For example, the carcinogenicity of chemical compounds increases in the following sequence:

Phenanthrene (γ = 1.40) < benz-1,2-anthracene (γ = 1.50) < dibenz-1,2,5,6-anthracene (γ = 1.57) < benz-3,4-pyrene (γ = 1.67) < dibenz-3,4,6,7-pyrene (γ = 1.71) < valen (γ = 2.29). (68)

At the same time, the molecular descriptor γ is obviously related to the amount of carcinogenic activity: an increase in the carcinogenicity of the substance is accompanied by an increase in the value of γ (as well as the sign of Z).

It is of fundamental importance that we do not use the idea of special local regions of molecules (that is, the K and L regions). On the contrary, we use molecular descriptors that characterize the properties of molecules in general. These descriptors make it possible to successfully interpret the observations that underlie the Pullmans model. Obviously, within the framework of our alternative model, it is possible to remove the following important remark made by Ladik J. : "... why just hydrocarbons that have joined the K region are carcinogenic, while the hydrocarbons that have joined in position L are inactive" [37]. From point of view of Pullmans model it is extremely puzzling [33] to observe that, whereas benz-3,4-pyrene is cancerogen, the 2′ and 3′ methyl derivatives are total inactive. However, the descriptors γ and Z for benz-3,4-pyrene are 1.67 and 2.88, and for methyl derivatives are 1.50 and 2.80, respectively. That is, the magnitude of the descriptors decreases. A similar decrease in descriptors occurs for inactive 6,7-dimethyl-3,4-benzphenanthrene (γ = 1.25, Z = 2.67) and 4,5-dimethylchrysene (γ = 1.13, Z = 2.59).

Some polycyclic hydrocarbons may have a high Z descriptor value, but they do not have carcinogenic activity. This may be due to the limiting factors of the manifestation of biological activity [8,9]. For example, the molecular descriptors Z, H and γ do not distinguish between isomer molecules. The benz-3,4-pyrene (Z = 2.875) is very active, and benz(e)pyrene (Z = 2.875) is weakly active. However, the affinity energy (A) for these molecules is markedly different. The affinity energy for benz(e)pyrene is equal to 0.07eV - 0.40 eV (comparison of different data). At the same time, the affinity energy for benz-3,4-pyrene, according to various sources, is in the range 0.70eV - 0.77 eV [36].

Pullman singled out [29], the following six polycyclic hydrocarbons, the carcinogenic whose activity significantly changes at little structural changes:

Benz-3,4-pyrene (Z = 2.875; A = 0.77 eV; ++++); dibenz-1,2,5,6-anthracene (Z = 2.833; A = 0.68 eV; ++); dibenz-1,2,5,6-phenanthrene (Z = 2.833; A = 0.31 eV; +); dibenz-1,2,3,4-phenanthrene (Z = 2.833; A = 0.40-0.43 eV; +); dibenz-1,2,7,8-anthracene (Z = 2.833; A = 0.23-0.69 eV; –/+); benz-1,2-pyrene (Z = 2.875; A = 0.49-0.61eV; –); dibenz-1,2,3,4-anthracene (Z = 2.833; A = 0.22-0.54 eV; –).

(69)

The experimental values of the affinity energy have been taken from the reference book [36]. To this series of compounds, the following polycyclic hydrocarbon can be added: dibenz-3,4,6,7-pyren (Z = 2.895, A = 0.82 eV; ++++). The lower bounds of the affinity energy indicate their parallelism with the carcinogenic activity of the molecules. It is possible that activity pyrene (Z = 2.846; +/–) is limited by the value of the affinity energy of the molecule ($A \geq 0.09$ eV).

The affinity energy gives a measure of the oxidative ability of a molecule. The transfer of an electron is usually accompanied by endoenergetic or exoenergetic processes that can affect the state of the objects with which the polycyclic hydrocarbon interacts. In the first case, for the transfer of an electron, energy is required to be at least two times greater than for benzo-3,4-pyrene. Since the carcinogen molecule interacts with molecular structures, in particular RNA and DNA [38], the locally released energy during electron transfer can be directed to the destruction of these ordered molecular structures. The concentration of hydrocarbons [30] significantly enhances the effect of the released energy in the interaction region.

The limiting effect on the carcinogenicity of a substance can provide the addition of an alkyl group to a polycyclic hydrocarbon [39]. First, this leads to a decrease in the molecular descriptors Z and γ. Secondly, an increase in the length of the alkyl group is accompanied by a change in the hydrophobic properties of the substance. According to Pullmans, steric hindrances can also be limiting factors. Shape, size and steric factors of the compounds are of importance for cancerogenic activity, but in the last resort this depends on the recipient tissue. It depends on the species, the particular strain of the test animal, its age, sex, its nutritional and hormonal state, and the phase of mitotic activity of a particular cell [4].

In accordance with the model of Pullmans, the chemical compounds LVIII and LIX (the numbering of Pullmans [33]) have a large spatial size and should be carcinogens. The same conclusion follows from calculations of molecular descriptors, which are equal to Z = 2.96 and 3.00, γ = 1.88 and 2.00, respectively. However, it was not possible to find the results of the experimental verification of these substances.

It is also useful to emphasize the high information content of the Z descriptor. For example, Table 18 shows the experimental values of the energy ΔE of the most intense electronic transition [40] for several aromatic hydrocarbons.

Table 18. The energies of the most intense electronic transition ΔE and the molecular descriptors of polyacenes

N	Chemical compound	ΔE, eV [40]	Z	H, *bits*	γ
1	Benzene	4.70	2.50	1.00	1.00
2	Naphthalene	3.94	2.67	0.99	1.25
3	Anthracene	3.27	2.75	0.98	1.40
4	Naphthacene	2.63	2.80	0.97	1.50
5	Pentacene	2.16	2.83	0.96	1.57

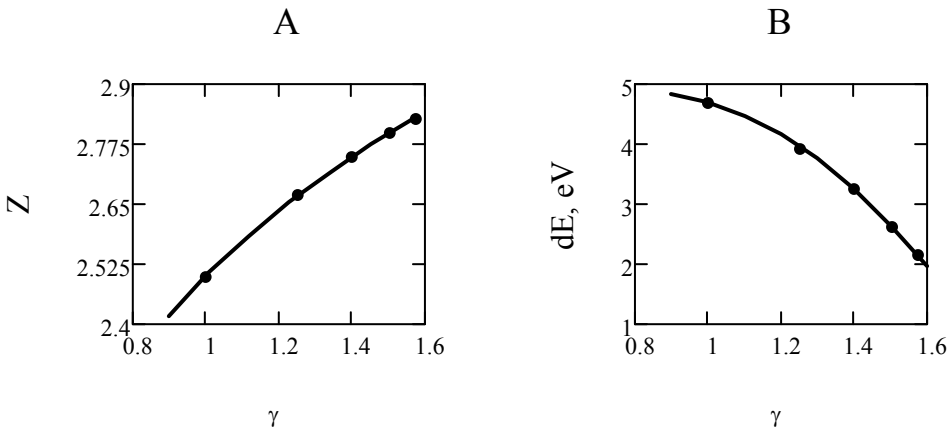

Fig. 9. (A) *The relationship between descriptors Z and γ. Statistics of the interrelation*: $Z(\gamma) = B + A \cdot \exp(- C \cdot \gamma)$, $B = 3.23 \pm 0.03$, $A = - 2.09 \pm 0.04$, $C = 1.05 \pm 0.06$, $RMSE = 0.0012$. (B) *The relationship between descriptors ΔE and γ. Statistics of the interrelation*: $\Delta E(\gamma) = A \cdot \gamma^2 + B \cdot \gamma + C$, $B = 7.85 \pm 0.99$, $A = - 4.78 \pm 0.39$, $C = 1.63 \pm 0.63$, $RMSE = 0.028$. *Here we use the following notation*: $dE \equiv \Delta E$.

Since the descriptor γ is closely related to the descriptor Z (Fig. 9A), the energy of the electronic transition ΔE also obviously correlates (Fig.9B) with the value of the molecular descriptor Z.

We can also note the high information content of the Z descriptor. This descriptor correlates with electron donor activity not only for polycyclic hydrocarbons, but also for other classes of compounds. The quadrupole splitting of the gamma-resonance line of tin in the Mössbauer spectrum of tin dibutyl chloride was studied [41]. The following order was obtained for a decrease in the donor properties of solvents towards to the organometallic compound $(C_4H_9)_2SnCl$ for seven solvents: CH_3SOCH_3 (Z = 2.60) > $HCON(CH_3)_2$ (Z = 2.50) > $[(CH_3)_2N]_3PO$ (Z = 2.35) > $CH_3OCH_2CH_2OCH_3$ (Z = 2.57) > tetrahydrofuran (Z = 2.30) > $H_5C_2OCH_2CH_2OC_2H_5$ (Z = 2.27) > $(C_2H_5)_2O$ (Z = 2.14). The donor ability reaches a minimum when the descriptor value Z for $(C_2H_5)_2O$ practically coincides with the value $Z = 2.17$ for the organometallic compound $(C_4H_9)_2SnCl$.

6. Conclusion

Classification rules allow us to identify the relationship between the biological response and the molecular structure of a chemical. The rules can be practically useful in the preliminary projection of the carcinogenic activity of new chemical compounds. It is important to emphasize that simple calculations of molecular descriptors require only knowledge of the chemical structural formula of a molecule. This approach makes it possible to considerably facilitate the search for new carcinogens, and also to draw the attention of researchers to poorly studied chemical compounds. However, it should be noted that the determination of the molecule descriptor is not sensitive to the study of iso-electronic molecular systems, and also when comparing the bioactivity of isomer molecules.

The ability of the Z descriptor to separate potentially carcinogenic compounds from non-carcinogenic substances is apparently not accidental but reflected the action of the real electrostatic molecular potential. This potential is generated by a set of charged particles (nuclei and electrons). The magnitude of the potential varies from molecule to molecule. It is very difficult to use the total molecular potential, which includes the Coulomb potential of nuclei and electrons. However, from analytic formulas for the pseudopotential [6] it follows that the general characteristic feature for the pseudopotential is the factor Z.

The change in the carcinogenic properties of molecules with a change in the molecular potential does not contradict the known notions of the mechanism of chemical carcinogenesis. The known data [26], as well as quantum-chemical calculations [27], allow us to conclude that, at least in a series of close congener

chemical compounds, their carcinogenicity is directly dependent on the ability to electrophylic attack.

It is suggested [26] that carcinogens induce DNA single-strand breaks. In this case, purine bases (especially guanine) are the target for them. In this regard, it should be noted that the molecular descriptor Z for all purine bases is larger than the threshold values (Table 6). The descriptors maximum values are achieved for guanine (Z = 3.50, H = 1.82*bits*). For other purine bases the value of the molecular descriptor Z is also higher than the threshold values: adenine (Z = 3.33), guanine (Z = 3.50), thymine (Z = 3.36), cytosine (Z = 3.23), uracil (Z = 3.5).

We used descriptors Z and H to study the radioprotective properties of sulfur-containing chemical compounds in the articles [7,8]. This opens the possibility to compare the preferred areas characterizing the radio-protective effect of sulfur-containing chemicals and their carcinogenic properties. Effective radioprotectors (dose ≤ 1 mM/kg; survival > 50%) are preferably in the following areas Z(protect.) ≤ 2.80 и H(protect.) ≤ 1.80*bits* [8]. At the same time, sulfur-containing carcinogens are likely to be in the following areas of Z(cancer) ≤ 3.15 and H(cancer) ≤ 1.87*bits*. Obviously, these areas for the descriptor of Z (or H) are overlap. Therefore, it is possible that effective radioprotectors may have carcinogenic properties. As well known from the literature there is the chemical compounds that have been tested for both antiradiation protection [28] and their carcinogenic activity [5]. For example, thiourea (Z = 3.00, H = 1.75*bits*) has radioprotective properties and at the same time is a carcinogen.

References

[1] Pliss G.B. In: Proceedings of the VIII International of Anti-Cancer Congress. 1963. vol. 2. Moscow. pp. 312-314.

[2] Haddow A. In: Proceedings of the VIII International of Anti-Cancer Congress. 1963. vol. 2. Moscow. pp.267-272

[3] Badger G.M. The Chemical Basis of Carcinogenic Activity. Charles C. Thomas – Publisher: Springfield-Illinois-USA. 1962.

[4] Schoental R. In: Clar E. "Polycyclic Hydrocarbons". V.1. Ch.18. London-N.Y., Berlin-Göttingen-Heidelberg. 1964.

[5] Carcinogenic substances. Handbook. Ed. Turusov V.S. (IARC Monographs on the Evaluation the Carcinonogenic Risk of Chemicals to Humans). Moscow. 1987.

[6] Mukhomorov V.K. "Simulation of Carcinogenic Activity of Polycyclic Hydrocarbons". In: Proceedings of the IV International Conference "Actual Problems of the Development of World Science". Part 1. pp. 86-97. Kiev. Ukraine. February 28, 2018. (in Russian).

[7] Mukhomorov V.K. *Adv. in Biological Chem*., 1, 2011, pp.1.

[8] Mukhomorov V.K. *Biomedical Statistics and Informatics*, 1 (2016) 24.

[9] Mukhomorov V.K. Modeling of Chemical Compounds Bioactivity. Relationships of Structure - Bioactivity. Lambert Academic Publisher. Saarbrűcken. Germany. 2012. (in Russian).

[10] Quastler G. In: Theory of Information in Biology. Ed. Blumenfeld LA. Moscow.1960.

[11] L'vovsky E.N. Statistical Methods for Constructing Empirical Formulas. Moscow. High School. 1988. (in Russian).

[12] Khalafyan A.A. Textbook. Statistica 6. Statistical Analysis of Data. Moscow. Publisher: Binomial. 2007. (in Russian).

[13] Fleiss J.L. Statistical Methods for Rates and Proportions. New York – Chichester – Brisbane – Toronto – Singapore. John Wiley & Sons, Inc. 1981.

[14] Forster E., Ronz B. Methoden der Korrelations- und Regressonanalise. Verlag Die Wirtschaft Berlin. 1979.

[15] Urbach V. Yu. Statistical Analysis in Biological and Medical Research. Moscow. Medicine. 1975. (in Russian).

[16] Pustylnik E.I. Statistical Methods of Analysis and Processing of Observations. Moscow. 1968. 288 p. (in Russian).

[17] Hollander M., Wolfe A. Nonparametric Statistical Methods. New York-London. John Wiley and Sons. Inc. 1973.

[18] http://potency.berkeley.edu/cpdb/html

[19] Rubenchik B.L. Biochemistry of Carcinogenesis. Kiev. 1977. (in Russian).

[20] Orris A., Yan Duuren B.L., Nelsen N. In: Proceedings of the VIII International of Anti-Cancer Congress. 1963. vol. 2, p. 305.

[21] Buu-Hoi N.P. *Cancer Res*., 24, 1964, pp.1511.

[22] Fukui J.A. et al. *Nippon Kaguki Zasshi*, 82, 1961, pp.474 (*Chem Abstr*. 57 (1962) 12378-b).

[23] Rubenchik B.L. Formation of Carcinogens from Nitrogen Compounds. Kiev. 1990.

[24] Leo A., Hansch C. *Chem. Rev*., 71, 1971, pp.525.

[25] Hidvedy T.Y., Arky I., Antoni F., Kõteles G. In: Proceedings of the VIII International of Anti-Cancer Congress. 1963. Moscow. vol. 2. p.225.

[26] Vilenchik MM Regularities of the Molecular-Genetic effect of Chemical Carcinogens. Moscow. 1977.

[27] Pullman B., Pullman A. Quantum Biochemistry. 1965.

[28] Doherty D. Radiation Protection and Recovery. New York. 1964.

[29] Pullman B. La Biochimie Electronique. Press Universitaires de France. 1963.

[30] Brookes P., Lawley P.D. *Nature*, 202, No.1964 4934, pp.781.

[31] Molecular Associations in Biology. Ed. by Pullman B. Academic Press, 1968. 217 p.
[32] Intermolecular Interactions: From Diatomics to Biopolymers. Ed. by Pullman B. John Wiley and Sons. Chichester-New-York. 1980.
[33] Pullman A., Pullman B. In: Advances in Cancer Research. Eds. by Greenstein J., Haddow A. Vol. 3. 1955. Acad. Press. Inc. Publisher. New-York.
[34] Badger G.M. The Chemical Basis of Carcinogenic Activity. Illinois, USA: Thomas-Publisher: Springfield, 1962. 98 p.
[35] Vylegzhanin N.I. *Oncology Affairs*. Vol. 21, Issue 5, 1952, pp.135. (in Russian).
[36] Gurvich L.V., Karachevtsev G.V., Kondratiev Yu. A. et al. The Energy of Chemical Linkages Crushing. Potentials of Ionization and Electron Affinity M. Nauka, 1974. (in Russian).
[37] Ladik J. Quantenbiochemie für Chemiker und Biologen. Akadimiai Kiado. Budapest. 1972.
[38] Heidelberger C., Moldenhauer M.G. *Cancer Research*, 16, 1956, pp.442.
[39] Fieser L.F., Putnam S.T. *J. Am. Chem. Soc*., 69. 1947. 1041.
[40] Peacock T.E. Electronic Properties of Aromatic and Heterocyclic Molecules. Acad. Press. London and New York. 1965.
[41] Goldanskii, V.I. *Angew. Chem.*, 79, 1967, pp.844.