ISSN 1070-3632, Russian Journal of General Chemistry, 2007, Vol. 77, No. 5, pp. 940–948. © Pleiades Publishing, Ltd., 2007.
Original Russian Text © I.G. Tsygankova, S.M. Zhenodarova, 2007, published in Zhurnal Obshchei Khimii, 2007, Vol. 77, No. 5, pp. 846–854.

# Quantitative Structure–Activity Relationship for Barbituric Acid Derivatives: Potential of the Fragment Approach

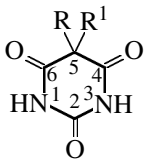
**I. G. Tsygankova and S. M. Zhenodarova**

*Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, ul. Institutskaya 3, Pushchino, Moscow oblast, 142290 Russia*
*e-mail: tsygan@iteb.ru*

Received August 22, 2006

**Abstract** — A previously suggested quantitative structure–property relationship was applied to estimation of the depressant action time of barbituric acid derivatives, using a set of 107 compounds as example. To choose molecular descriptors and decrease the number of parameters of the correlation equation, the methods of projection onto latent structures and step-by-step regression were used. In both cases, the calculated values are fairly well consistent with the experimental data, but the method of step-by-step regression allows simple interpretation of the correlation equation and determination of structural elements important for manifestation of the biological activity.
**DOI:** 10.1134/S1070363207050222

Barbiturates are barbituric acid derivatives in which the hydrogen atoms in position 5 are replaced with alkyl, cycloalkyl, or aromatic radicals. These compounds are widely used for a long time in medicine as soporifics, anticonvulsants, and narcotics.



However, the researchers' interest in these compounds, which can cause all levels of depression of the central nervous system, from mild sedative effect and hypnosis to deep coma and death, remains steadily high, because the mechanisms of their neurochemical action remain poorly understood.

As early as 1967 [1], it was concluded that the biological activity of barbiturates is mainly determined by their relative hydrophobicity, and the steric and electronic effects play an insignificant role, although some differences in the properties of barbiturates could not be attributed solely to the difference in the lipophilicity. Therefore, Hsieh and Dorsey [2] suggested that these differences could be associated with differences in the absorption, distribution, metabolism, and, finally, in the specific interaction of barbiturates with the corresponding receptors. Barbiturates are still fairly frequently used as a convenient model in various studies: for examining the distribution of drugs in lipid bilayers and biological membranes [2–5], in blood and various organs or tissues of the whole body [6], or, e.g., as molecular probes for noninvasive imaging of matrix metalloproteinases in diagnostics of inflammatory, tumor, and cardiovascular diseases [7]. As a rule, these studies were aimed to establish a correlation between the chromatographic characteristics (retention indices, capacity factors, etc.) and biological activity and structure of barbiturates [1–5, 8, 9], but the molecular descriptors chosen (distribution ratio between organic and aqueous phases, dipole moment or dielectric permittivity of mobile phase, molecular refraction, molar extinction coefficient in UV spectrum, etc.) do not show a simple correlation with the molecular structure, so that the suggested correlation equations do not allow estimation of the property with variation of the structure, or prediction of the structure of a molecule with a preset activity.

In this study we suggest to use the fragment presentation of the molecular structure for studying the structure–activity relationships for barbituric acid derivatives. Fragment descriptors of the molecular structure are pictorial and have a clear structural sense, but constructing a quantitative structure–property relationship using these descriptors is, as a rule, difficult because of the great number of possible fragment contributions at a limited set of the experimental data. The goal of this study is to compare different regres-

sion approaches (step-by-step regression, projection onto latent structures) in construction of correlation equations using fragment molecular descriptors and to examine the possibility of physical interpretation of the correlation equations obtained. As the measure of the activity of the barbiturates we will use their depressant action time (DAT). The DAT data are available for a large set of barbiturates [10]. Unfortunately, these data were taken from different sources and, as indicated in [10], were determined in the original studies under different conditions. Nevertheless, assuming that all the processes determining the drug effect are associated with intermolecular interactions and therefore depend on the compound structure, we will construct the general structure–DAT correlation for the whole set of the available experimental data (Table 1).

Previously [11, 12] we suggested a correlation equation for estimating various physicochemical properties of organic compounds of different classes, in which a property is represented as a sum of contributions from fragments constituting the molecule:

$$P = \Sigma P_i + \Sigma\Sigma P_{i,j} + \Sigma\Sigma\Sigma P_{i,j,k} + \ldots,$$

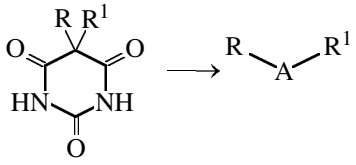
where $P_i$ is the contribution to a property of one fragment of number $i$; $P_{i,j}$ and $P_{i,j,k}$ are multifragment contributions to a property, determined by interaction of two and more fragments.

This approach makes it possible to choose a representation of molecules by fragments of different sizes, corresponding to one atom or a group of atoms. Later it was shown that application of this approach to various kinds of biological activity allows its quantitative estimation with a reasonable accuracy [13, 14]. An important generalization of the fragment approach is that the correlation equation can include not only one-fragment contributions but also multifragment contributions reflecting the mutual influence (interaction) of the fragments, which depends on the nature of fragments and distance between them. In this study, we will restrict the consideration to pair interactions and will determine the distance $r$ between the fragments in the topological sense, i.e., we will take the distance $r$ equal to the number of bonds in the molecular skeleton along the shortest path between the fragments. Thus, the expression for a property will take the form

$$P \approx \text{Const} + \Sigma n_i P_t + \Sigma k_{t,m,r} P_{t,m,r}. \qquad (1)$$

where $n_t$ is the number of fragments bonded to atoms of a definite kind $t$; $k_{t,m,r}$, number of pairs of fragments of types $t$ and $m$ located at a definite distance $r$. The fragment contributions $P_t$ and $P_{t,m,x}$ are parameters of a correlation equation, determined from the experimentally measured activities. The coefficients $n_t$ and $k_{t,m,x}$ are quantitative descriptors of the molecular structure.

To describe the structure of barbiturates, we will consider barbituric acid without substituents R and $R^1$ (pyrimidinetrione ring) as a single fragment A:




We will describe alkyl or alkenyl substituents R and $R^1$ in position 5 by fragments bonded to the carbon atoms. Saturated carbon atoms with any number of linked hydrogen atoms can be naturally considered as fragments of one type C. Fragments of another type C* correspond to C atoms involved in a double bond. The substituents in position 5 are symmetry-related. Therefore, we will consider the fragments of the first and second susbtituents to be identical. It is known that terminal carbon atoms of $CH_3$ groups differ in the behavior from internal C atoms [15, 16]. They are more mobile and have larger accessible surface; apparently, they interact with atoms of other molecules differently, compared to the other atoms. This fact can be taken into account in a correlation equation by assigning fragments bound to terminal $CH_3$ groups to a special type T. Thus, to describe the structure of barbiturate molecules, we will use four types of fragments: A, C, C*, and T.

The choice of fragment types is largely based on the intuition and is limited to the set of experimental data. The number of possible pairs of fragments of four types is ten. It should be noted, however, that fragment A (pyrimidinetrione ring) essentially differs from fragments C, C*, and T. Being bulky, fragment A can perturb the interaction of fragments C, C*, and T belonging to different substituents. Therefore, the interaction of fragments belonging to different substituents should be assumed to be different from the inetraction of fragments of the same substituent. For example, pairs of saturated C atoms in radical R or $R^1$ will be characterized by CC type of interaction, and pairs of C atoms of different substituents, by another type, CC1. Then our model will include 16 types of interactions. The longest distance in the set of compounds under consideration is 10. Thus, the total number of descriptors (i.e., types of one- and two-fragment contributions) in the model is $4 + (16 \times 10) = 164$. After excluding zero descriptors, descriptors with the variance less than 0.1, and strongly correlated

**Table 1.** Depressant action time (DAT) of barbiturates considered in this study [10]

| Run no. | Barbiturate | | DAT, min | Run no. | Barbiturate | | DAT, min |
|---|---|---|---|---|---|---|---|
| | R | $R^1$ | | | R | $R^1$ | |
| 1 | $CH_3$ | $(CH_3)_3CCH_2$ | 580 | 55 | $CH_3(CH_2)_2$ | $CH_3(CH_2)_4$ | 4 |
| 2 | 〃 | $CH_3(CH_2)_5$ | 260 | 56 | 〃 | $CH_3CH_2CH(CH_3)CH_2$ | 165 |
| 3 | 〃 | $CH_3(CH_2)_3CH(CH_3)$ | 227 | 57 | 〃 | $CH_3(CH_2)_5$ | 1 |
| 4 | 〃 | $CH_3(CH_2)_3CH(CH_3CH_2)CH_2$ | 223 | 58 | 〃 | $CH_3(CH_2)_6$ | 15 |
| 5 | 〃 | $H_2C=C(CH_3)$ | 60 | 59 | 〃 | $CH_3HC=CH$ | 60 |
| 6 | 〃 | $CH_3CH=C(CH_3)$ | 120 | 60 | 〃 | $CH_3CH_2HC=CH$ | 18 |
| 7 | 〃 | $CH_3CH_2HC=C(CH_3)$ | 60 | 61 | 〃 | $(CH_3)_2CHHC=CH$ | 18 |
| 8 | 〃 | $CH_3HC=C(CH_3CH_2)$ | 60 | 62 | 〃 | $H_2C=C(CH_3)$ | 168 |
| 9 | 〃 | $CH_3(CH_2)_2HC=C(CH_3)$ | 60 | 63 | 〃 | $CH_3HC=C(CH_3)$ | 30 |
| 10 | 〃 | $(CH_3)_2CHC=C(CH_3)$ | 36 | 64 | 〃 | $CH_3CH_2HC=C(CH_3)$ | 18 |
| 11 | 〃 | $CH_3(CH_2)_3HC=C(CH_3)$ | 24 | 65 | 〃 | $CH_3HC=C(CH_3CH_2)$ | 24 |
| 12 | $CH_3CH_2$ | $CH_3CH_2$ | 1400 | 66 | 〃 | $H_2C=CHCH(CH_3)$ | 420 |
| 13 | 〃 | $CH_3(CH_2)_2$ | 1140 | 67 | 〃 | $H_2C=C(CH_3)CH_2$ | 300 |
| 14 | 〃 | $CH_3CH(CH_3)$ | 1520 | 68 | 〃 | $CH_3HC=CHCH_2$ | 120 |
| 15 | 〃 | $CH_3(CH_2)_3$ | 450 | 69 | $(CH_3)_2CH$ | $(CH_3)_2CHCH_2$ | 25 |
| 16 | 〃 | $CH_3CH(CH_3)CH_2$ | 540 | 70 | 〃 | $CH_3HC=CH$ | 36 |
| 17 | 〃 | $CH_3CH_2CH(CH_3)$ | 600 | 71 | 〃 | $CH_3CH_2HC=CH$ | 36 |
| 18 | 〃 | $CH_3(CH_2)_4$ | 220 | 72 | 〃 | $CH_3(CH_2)_2HC=CH$ | 18 |
| 19 | 〃 | $CH_3CH_2CH(CH_3)CH_2$ | 190 | 73 | 〃 | $(CH_3)_2CHHC=CH$ | 12 |
| 20 | 〃 | $(CH_3)_3CCH_2$ | 200 | 74 | 〃 | $CH_3CH_2HC=C(CH_3)$ | 18 |
| 21 | 〃 | $CH_3CH_2CH(_{CH_3CH_2})$ | 300 | 75 | 〃 | $CH_3C=C(CH_3CH_2)$ | 18 |
| 22 | 〃 | $CH_3(CH_2)_5$ | 45 | 76 | 〃 | $H_2C=CHCH(CH_3)$ | 210 |
| 23 | 〃 | $CH_3(CH_2)_2CH(CH_3)CH_2$ | 210 | 77 | 〃 | $CH_3HC=CHCH_2$ | 200 |
| 24 | 〃 | $CH_3CH_2C(CH_3)_2CH_2$ | 60 | 78 | $CH_3(CH_2)_3$ | $CH_3CH_2CH(CH_3)$ | 16 |
| 25 | 〃 | $CH_3(CH_2)_3CH(CH_3)$ | 90 | 79 | 〃 | $(CH_3)_3C$ | 1 |
| 26 | 〃 | $CH_3CH_2CH(_{CH_3CH_2)CH_2}$ | 300 | 80 | 〃 | $CH_3HC=CH$ | 12 |
| 27 | 〃 | $CH_3(CH_2)_6$ | 120 | 81 | 〃 | $CH_3CH_2HC=CH$ | 18 |
| 28 | 〃 | $(CH_3)_2CHCH_2CH(CH_3)CH_2$ | 54 | 82 | 〃 | $H_2C=C(CH_3)$ | 90 |
| 29 | 〃 | $(CH_3)_2CH(CH_2)_2CH(CH_3)$ | 50 | 83 | 〃 | $CH_3HC=C(CH_3)$ | 60 |
| 30 | 〃 | $CH_3CH_2CH(CH_3)CH_2CH(CH_3)$ | 74 | 84 | 〃 | $H_2C=CHCH(CH_3)$ | 110 |
| 31 | 〃 | $CH_3(CH_2)_2CH(CH_3CH_2CH_2)$ | 81 | 85 | 〃 | $CH_3HC=CHCH_2$ | 40 |
| 32 | 〃 | $CH_3(CH_2)_2CH(CH_3)CH_2CH_2CH_2$ | 60 | 86 | 〃 | $(CH_3)_2C=CHCH_2$ | 30 |
| 33 | 〃 | $CH_3CH_2CH(CH_3)CH_2CH(CH_3)CH_2$ | 60 | 87 | $H_2C=CH$ | $CH_3CH_2CH_2CH_2$ | 288 |
| 34 | 〃 | $CH_3(CH_2)_3CH(CH_3CH_2)CH_2$ | 75 | 88 | 〃 | $(CH_3)_3CCH_2$ | 192 |
| 35 | 〃 | $CH_3(CH_2)_4CH(CH_3CH_2)$ | 60 | 89 | $H_2C=C(CH_3)$ | $H_2C=CHCH_2$ | 102 |
| 36 | 〃 | $CH_3(CH_2)_5CH(CH_3)$ | 150 | 90 | 〃 | $(CH_3)_2CHCH_2$ | 90 |
| 37 | 〃 | $CH_3(CH_2)_2CH(CH_3)CH(CH_3CH_2)CH_2$ | 240 | 91 | 〃 | $CH_3(CH_2)_4$ | 30 |
| 38 | 〃 | $(CH_3)_2CH(CH_2)_2CH(CH_3CH_2)CH_2$ | 120 | 92 | 〃 | $(CH_3)_3CCH_2$ | 18 |
| 39 | 〃 | $H_2C=CH$ | 288 | 93 | $CH_3HC=C(CH_3)$ | $H_2C=CHCH_2$ | 30 |
| 40 | 〃 | $H_2C=C(CH_3)$ | 150 | 94 | $H_2C=CHCH_2$ | $CH_3(CH_2)_3CH(CH_3)$ | 108 |
| 41 | 〃 | $CH_3CH_2HC=CH$ | 18 | 95 | 〃 | $H_2C=CHCH(CH_3)$ | 456 |
| 42 | 〃 | $CH_3HC=C(CH_3)$ | 180 | 96 | 〃 | $H_2C=C(CH_3)CH_2$ | 380 |
| 43 | 〃 | $(CH_3)_2C=CH$ | 240 | 97 | 〃 | $(CH_3)_2CHCH_2$ | 162 |
| 44 | 〃 | $CH_3(CH_2)_2HC=CH$ | 96 | 98 | 〃 | $(CH_3)_3CCH_2$ | 96 |
| 45 | 〃 | $CH_3CH_2HC=C(CH_3)$ | 24 | 99 | 〃 | $H_2C=CHCH_2$ | 880 |
| 46 | 〃 | $(CH_3)_2CHHC=CH$ | 12 | 100 | 〃 | $(CH_3)_2CH$ | 720 |
| 47 | 〃 | $CH_3HC=C(CH_3CH_2)$ | 42 | 101 | 〃 | $CH_3(CH_2)_2CH(CH_3)$ | 150 |
| 48 | 〃 | $CH_3(CH_2)_2HC=C(CH_3)$ | 72 | 102 | $CH_3HC=CHCH_2$ | $(CH_3)_3CCH_2$ | 40 |

**Table 1.** (Contd.)

| Run no. | Barbiturate | | DAT, min | Run no. | Barbiturate | | DAT, min |
|---|---|---|---|---|---|---|---|
| | R | $R^1$ | | | R | $R^1$ | |
| 49 | " | $CH_3(CH_2)_3HC=C(CH_3)$ | 6 | 103 | " | $CH_3(CH_2)_2CH(CH_3)$ | 66 |
| 50 | " | $CH_3CH_2HC=C(CH_3CH_2CH_2)$ | 6 | 104 | " | $CH_3CH_2CH(CH_3)$ | 120 |
| 51 | " | $H_2C=CHCH(CH_3)$ | 720 | 105 | " | $(CH_3)_2CHCH_2$ | 45 |
| 52 | " | $H_2C=C(CH_3)CH_2$ | 326 | 106 | $(CH_3)_2C=CHCH_2$ | $(CH_3)_2C=CHCH_2$ | 70 |
| 53 | " | $CH_3HC=CHCH_2$ | 372 | 107 | " | $CH_3CH_2CH(CH_3)$ | 120 |
| 54 | " | $CH_3(CH_2)_2CH(CH_3)$ | 180 | | | | |

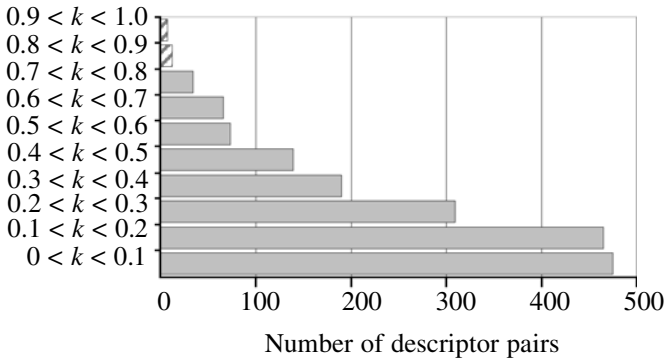
(correlation coefficient >0.98) descriptors, 60 descriptors remain in the model. The set under consideration consists of 107 barbituric acid derivatives. In the general matrix including all the descriptors of our model, the number of columns (60 unknowns) is smaller than the number of number of rows (107 experimental points), which allows the use of multiple regression methods, i.e., determination of parameters of Eq. (1), e.g., by the least-squares method. To elucidate the role of descriptors, it is important that they should not mask each other, i.e., should be independent to the maximum possible extent. In the mathematical sense, this means that the columns in the general matrices should be orthogonal. Fragment descriptors are not orthogonal by construction. To convert the matrix of descriptors into an orthogonal matrix, we used the method of projection onto latent structures (PLS), which is widely used in constructing structure–property correlations [17].

Since the experimental times of depressant effect of the set of compounds under consideration vary in a range covering three orders of magnitude (Table 1), we presented them in the logarithmic form.

Analysis of the experimental times of depressant effect of barbiturates allows certain general conclusions: (1) an increase in the total number of carbon atoms in saturated substituents decreases DAT of barbiturates; (2) branching of the carbon chain, as a rule, makes DAT longer compared to the linear analogs; and (3) the presence of a double bond in the side substituent makes DAT shorter. However, certain deviations from these trends can be noted. Some examples are given in Table 2. The upper part of the table over the triple line includes sets of compounds with increasing chain length in substituents, and the lower part, pars of isomers with linear and branched substituents. It is seen that (1) an increase in the length of the linear substituent from $C_6$ to $C_7$ increases DAT, in contrast to a decrease in DAT in going from $C_2$ to $C_6$ (nos. 12, 13, 15, 18, 22, 27; nos. 14, 17, 25, 36); (2) an increase in the size of a substituent containing an unsaturated bond can exert an insignificant effect on DAT (nos. 39, 87); (3) in the derivatives with $R^1$ = $C_5$, introduction of the $CH_3$ group at the first or second C atom makes DAT somewhat shorter (nos. 18 and 54, 18 and 19), whereas in the other cases DAT increases (nos. 13 and 14, 15 and 17, 18 and 21, 22 and 23). It is interesting how the revealed trends in the dependence of DAT on the structure of barbiturates will be reflected in the fragment model of the structure–activity correlation.

The calculated molecular descriptors for barbituric acid derivatives preserve a high degree of mutual correlation even after the pretreatment. The figure shows a histogram of the distribution of the pair correlation coefficients for the whole set of descriptors, from which it follows that a noticeable number of descriptor pairs have the correlation coefficient higher than 0.8 (the corresponding rectangles are hatched with oblique lines). To select the descriptors and decrease the number of parameters of the correlation equation, we used two approaches: projection onto latent structures [17] and step-by-step regression [18].

In the first case, a descriptor matrix was converted by a linear transformation into an orthogonal matrix,




Distribution of the pair correlation coefficients ($k$) of structural descriptors.

**Table 2.** Variation of DAT of barbiturates with the structure of the substituents

| Compound[a] | R | R' | ln (DAT) | |
|---|---|---|---|---|
| | | | experiment [10][b] | calculation[c] |
| 12 | $CH_3CH_2$ | $CH_3CH_2$ | 7.24 | 7.36 |
| 13 | " | $CH_3(CH_2)_2$ | 7.04 | 6.11 |
| 15 | " | $CH_3(CH_2)_3$ | 6.11 | 5.17 |
| 18 | " | $CH_3(CH_2)_4$ | 5.39 | 4.79 |
| 22 | " | $CH_3(CH_2)_5$ | 3.81 | 3.86 |
| 27 | " | $CH_3(CH_2)_6$ | 4.79 | 4.50 |
| 14 | $CH_3CH_2$ | $CH_3CH(CH_3)$ | 7.33 | 6.54 |
| 17 | " | $CH_3CH_2CH(CH_3)$ | 6.40 | 5.97 |
| 25 | " | $CH_3(CH_2)_3CH(CH_3)$ | 4.50 | 4.65 |
| 36 | " | $CH_3(CH_2)_5CH(CH_3)$ | 5.01 | 4.36 |
| 39 | $H_2C=CH$ | $CH_3CH_2$ | 5.66 | 5.65 |
| 87 | " | $CH_3(CH_2)_3$ | 5.66 | 5.12 |
| 13 | $CH_3CH_2$ | $CH_3(CH_2)_2$ | 7.04 | 6.11 |
| 14 | " | $CH_3CH(CH_3)$ | 7.33 | 6.54 |
| 15 | $CH_3CH_2$ | $CH_3(CH_2)_3$ | 6.11 | 5.17 |
| 17 | " | $CH_3CH_2CH(CH_3)$ | 6.40 | 5.97 |
| 18 | $CH_3CH_2$ | $CH_3(CH_2)_4$ | 5.39 | 4.79 |
| 54 | " | $CH_3(CH_2)_2CH(CH_3)$ | 5.19 | 5.02 |
| 18 | $CH_3CH_2$ | $CH_3(CH_2)_4$ | 5.39 | 4.79 |
| 19 | " | $CH_3CH_2CH(CH_3)CH_2$ | 5.25 | 5.32 |
| 18 | $CH_3CH_2$ | $CH_3(CH_2)_4$ | 5.39 | 4.79 |
| 21 | " | $CH_3CH_2CH(CH_3CH_2)$ | 5.70 | 5.50 |
| 22 | $CH_3CH_2$ | $CH_3(CH_2)_5$ | 3.81 | 3.86 |
| 23 | " | $CH_3(CH_2)_2CH(CH_3)CH_2$ | 5.35 | 4.95 |

[a] Here and in Tables 4–6, the compound numbering is the same as in Table 1. [b] For the sake of convenience, the experimental values are converted to the logarithmic scale. [c] Calculated by Eq. (1) with the parameters listed in Table 3.

taking into account the requirement of minimal remainders, and a correlation equation was constructed with the orthogonal components obtained [19]. The DAT estimates obtained with this equation have a standard deviation from the experimental values of about 0.65 and the correlation coefficient of 0.8816, with seven parameters retained in the correlation equation; the values of these parameters are 6.87, 3.52, 3.10, 2.87, 1.70, 1.30, and 0.96, respectively. When checking the equation by the leave-one-out method, the correlation coefficient of the control is 0.7163 and the rms deviation of the control is 0.99 log unit. Thus, using the equation obtained, we can calculate DAT of barbituric acid derivatives with a reasonable accuracy. However, to detect a trend in the property variation and to understand how the structure should be altered to attain the required change in the property, it is necessary to reveal the role of definite structural elements. For this purpose, the correlation equation obtained should be returned from orthogonal components to structural descriptors, taking into account weight coefficients. In so doing, the equation will again include the whole set of descriptors with their mutual dependences and the coefficients will become fairly close in the absolute value, which does not allow us to distinguish a small number of the most

important contributions. Thus, the method of projection onto latent structures, being very useful in estimation of the activity, does not allow us to physically interpret the correlation equation and to suggest a meaningful model of the structure–activity corelation.

In the method of step-by-step regression [18], each step involves inclusion in the correlation equation of a descriptor that is the most strongly correlated with the descriptors remaining after inclusion or exclusion of a desciptor in the previous step. The decision on inclusion or exclusion of a descriptor is made on the basis of calculating the partial *F* test for each variable [18]. The result of the selection depends on the choice of the initial conditions. The step-by-step regression allowed us to obtain several close sets of descriptors. The characteristics of an equation constructed with one of such sets of descriptors are given in Table 3. This equation, compared to equations based on other sets of descriptors, ensures the lowest rms deviation for the DAT estimates.

In analysis of the model, let us assume for definiteness sake that the contributions increasing DAT are favorable for the property, although in some cases it may be necessary to find a drug with the excretion as fast as posible, and in this case compounds with short DAT would be preferable. Some of the selected descriptors have direct physical interpretation: C* is the number of unsaturated carbon atoms, i.e., doubled number of double bonds; TT-2 characterizes the number of methyl groups at the next to last C atom of substituent R or $R^1$; AC*-3 in our set of compounds characterizes the number of double bonds separated by one C atom from the pyrimidinetrione ring. The coefficients at these contributions suggest that, the larger the numbers of double bonds in substituents and of short branchings at the next to last C atom, the shorter the DAT. Also, the separation of the double bond from fragment A by one atom is more preferable for increasing DAT than the direct neighborhood, other things being the same.

Interpretation of the other contributions is not so straightforward. The CC*-3 contribution is present in all the barbituric acid derivatives in which one substitunet is unsaturated and the other contains no less than two carbon atoms. The positive sign of the coefficient at this contribution in the correlation equation suggests that specifically the absence of this contributions in the methyl derivatives (R = $CH_3$) is responsible for the shorter DAT of these compounds, compared, e.g., to the derivatives with R = ethyl (cf., e.g., compound nos. 5 and 40 in Table 1). The CC*-1 and CC*-3 contributions have different signs and are present in unequal amounts only in compounds with chains of

**Table 3.** Parameters of equation obtained by step-by-step regression

| Contribution | Coefficient |
|---|---|
| Const | 8.09 ± 0.77 |
| CC1-6 | –1.88 ± 0.58 |
| C*C-1 | –1.64 ± 1.04 |
| C* | –1.60 ± 0.40 |
| CC-1 | –1.26 ± 0.42 |
| TC1-2 | –0.49 ± 0.15 |
| TT-2 | –0.45 ± 0.19 |
| TC1-3 | –0.37 ± 0.24 |
| TC1-6 | –0.31 ± 0.21 |
| TC1-4 | –0.29 ± 0.17 |
| TC-2 | 0.23 ± 0.19 |
| CC*1-3 | 0.75 ± 0.35 |
| CC-4 | 0.89 ± 0.53 |
| CC-2 | 1.01 ± 0.47 |
| CC*-2 | 1.02 ± 0.74 |
| AC*-3 | 2.21 ± 0.50 |
| *r* | 0.8846 |
| *s* | 0.65 |
| Fisher test | 21.83 |

three and more C atoms connected to unsaturated C atoms. The difference of these contributions, and also the TC-2 contribution provide longer DAT for compounds with the propyl substituent; however, the compounds present in the experimental set are insufficient for obvious demonstration of the contribution of these descriptors. The contributions of the descriptors CC1-6, TC1-6, TC1-5, TC1-4, and TC1-3 are all negative, i.e., an increase in the substituent chain length to six carbon atoms results in shorter DAT. As the chain length is increased further, the total contribution of these descriptors remains unchanged. However, the activity changes owing to an increase by unity in the number of CC-4, CC-2, and CC-1 contributions. Their total contribution is 0.884 + 1.008 – 1.26 = 0.632 log unit, i.e., the DAT increases as the length of the linear part of the chain is increased over six atoms (Table 2, compound nos. 12, 13, 15, 18, 22, 27). The suggested model, though explaining some trends in the dependence of DAT of barbiturates on the molecular structure, remains approximate. This is seen from Table 2: The calculated DAT values do not reflect the above-discussed features (compound nos. 14, 17, 25, 36, and also 39 and 87, 18 and 54, 18 and 19). Probably, retention of three-fragment and more complex contributions in Eq. (1) would improve the accuracy of the model.

Published data are also available for some other

**Table 4.** Hypnotic activity of barbiturates

| Barbiturate, compound no. in Table 1 | | log (1/C) [b] | | | |
|---|---|---|---|---|---|
| R [a] | R1 [a] | **I** | | **II** | |
| | | experiment [9] | calculation [c] | experiment [1] | calculation [c] |
| | 12 | 2.91 | 3.13 | 3.09 | 3.33 |
| | 14 | 3.31 | 3.33 | 3.30 | 3.33 |
| | 15 | 3.53 | 3.24 | 3.74 | 3.52 |
| | 22 | 3.08 | 3.24 | | |
| | 54 | 4.05 | 4.05 | | |
| | 97 | 3.63 | 3.73 | | |
| | 99 | 3.54 | 3.49 | | |
| | 100 | 3.60 | 3.51 | | |
| | 101 | 4.02 | 3.92 | | |
| $CH_3CH_2$ | $(CH_3)_2CH(CH_2)_2$ | 3.59 | 3.55 | 3.75 | 3.71 |
| $CH_2=CHCH_2$ | $CH_3(CH_2)_2$ | 3.47 | 3.42 | | |
| $CH_2=CHCH_2$ | $(CH_3)_2CH(CH_2)_2$ | 3.49 | 3.73 | | |
| $(CH_3)_2CH$ | $CH_3(CH_2)_3$ | 3.47 | 3.44 | | |
| $CH_3(CH_2)_3$ | $CH_3(CH_2)_3$ | 3.45 | 3.35 | 2.84 | 2.84 |
| $CH_3CH_2$ | $(CH_3)_2CHCH_2$ | | | 3.63 | 3.71 |
| $CH_3(CH_2)_2$ | $CH_3(CH_2)_2$ | | | 3.55 | 3.52 |
| $CH_3(CH_2)_2$ | $(CH_3)_2CH$ | | | 3.63 | 3.71 |
| $CH_3(CH_2)_2$ | $(CH_3)_2CHCH_2$ | | | 3.48 | 3.33 |
| | *r* | | 0.8355 | | 0.9321 |
| | *s* | | 0.14 | | 0.12 |
| | Fisher test | | 6.93 | | 13.24 |

[a] Given for barbiturates not included in Table 1. [b] Minimal hypnotic dose (mol $kg^{-1}$) for rats (**I**) and rabbits (**II**). [c] In calculation by Eq. (1), we used the C*, TA-2, TC-2 (for **I**) and TT-4, CC1-6, and TT1-6 (for **II**) descriptors, respectively.

kinds of activity of barbituric acid derivatives. Using the fragment representation of the molecule, it is possible to obtain simple (including 2–4 descriptors) descriptive models for these kinds of activity (Tables 4, 5). The estimates of these kinds of activity are reasonably consistent with the experimental data, but the predictive power of these models is low because the available data are few.

**Table 5.** Inhibition by barbiturates of the fission of Arbacia egg cells and of the oxygen uptake by rat brain in vivo [1]

| Barbiturate, compound no. in Table 1 | | log (1/C) [b] | | | |
|---|---|---|---|---|---|
| R [a] | R1 [a] | **I** | | **II** | |
| | | experiment | calculation [c] | experiment | calculation [c] |
| | 12 | 1.49 | 1.78 | 1.32 | 1.44 |
| | 14 | 1.79 | 1.78 | 1.89 | 1.97 |
| | 15 | 2.40 | 2.93 | 2.80 | 2.51 |
| | 18 | 2.82 | 2.80 | | |
| | 21 | 2.85 | 2.75 | | |
| | 22 | 3.12 | 3.07 | 3.40 | 3.59 |
| | 34 | 3.70 | 3.78 | | |
| | 54 | 2.92 | 2.75 | 3.07 | 3.05 |

**Table 5.** (Contd.)

| Barbiturate, compound no. in Table 1 | | log (1/C) [b] | | | |
|---|---|---|---|---|---|
| R [a] | R1 [a] | **I** | | **II** | |
| | | experiment | calculation [c] | experiment | calculation [c] |
| | 97 | 2.41 | 2.47 | 2.80 | 2.71 |
| | 99 | 1.79 | 1.78 | | |
| | 100 | 2.01 | 1.78 | 2.15 [3] | 2.18 |
| | 101 | 3.62 | 3.59 | 3.19 | 3.25 |
| $CH_3CH_2$ | $(CH_3)_2CH(CH_2)_2$ | 2.82 | 2.75 | 3.12 | 3.05 |
| $CH_3CH_2$ | $CH_3CH{=}CHCH(CH_3)$ | 3.30 | 3.30 | | |
| $CH_3CH_2$ | $(CH_3)_2CHCH_2CH(CH_3)$ | 2.82 | 2.97 | | |
| | *r* | | 0.9812 | | 0.9798 |
| | *s* | | 0.13 | | 0.13 |
| | Fisher test | | 64.60 | | 71.84 |

[a] Given for barbiturates not included in Table 1. [b] (*C*) Concentration (M) causing 50% inhibition of the fission of Arbacia egg cells (**I**) and of the oxygen uptake by rat brain in vivo (**II**). [c] In calculation by Eq. (1), we used the C*, TC*-2, TC-3, C1C*-5 (for **I**) and C, C* (for **II**) descriptors, respectively.

In a widely used approach, the biological activity is estimated from the distribution ratio between the organic and aqueous phases ($\log P$). With $\log P$ as molecular descriptor, the hydrophobic and polar properties of compounds are taken into account, and good models are obtained. However, as a rule, the reliable experimental values of $\log P$ are insufficient, and various estimates are used instead [1]. The fragment approach proved to be efficient in calculations of $\log P$ [20–22], but the calculation of $\log P$ means an additional intermediate step introducing additional inaccuracy into the model. The data for 15 compounds from the set under consideration, for which data on $\log P$ obtained by a similar procedure are available [1], are given in Table 6. The DAT values of barbituric acid derivatives estimated using the fragment approach are no less accurate than those estimated through $\log P$.

Thus, using the fragment descriptors of the molecular structure, we constructed a model for the correlation of the structure of barbituric acid derivatives with the time of their depressant action. The model gives reasonably accurate estimates of the action time for a set of 107 derivatives. Despite diversity of the experimental data, they can be described by a common equation. Orthogonalization of the descriptor matrix by the method of projection onto latent structures allows the number of equation parameters to be decreased to seven, with the reasonable accuracy of the action time estimates preserved. A decrease in the number of descriptors by the step-by-step regression method leaves 15 parameters in the equation ensuring

**Table 6.** Action time of barbiturates, calculated from $\log P$ and using the fragment model

| Compound | ln (DAT) | | |
|---|---|---|---|
| | experiment [10] | calculation | |
| | | Const + $b^*$ $\log P$ | by Eq. (1) [a] |
| 12 | 7.24 | 7.05 | 7.36 |
| 13 | 7.04 | 6.81 [b] | 6.11 |
| 14 | 7.33 | 6.73 | 6.54 |
| 15 | 6.11 | 5.97 | 5.17 |
| 17 | 6.40 | 6.19 | 5.97 |
| 18 | 5.39 | 5.43 | 4.79 |
| 21 | 5.70 | 5.65 | 5.50 |
| 22 | 3.81 | 4.90 | 3.86 |
| 27 | 4.79 | 3.83 [b] | 4.50 |
| 34 | 4.32 | 4.04 | 4.75 |
| 54 | 5.19 | 5.65 | 5.02 |
| 97 | 5.09 | 5.97 | 5.40 |
| 99 | 6.78 | 6.62 | 5.32 |
| 100 | 6.58 | 6.51 | 6.06 |
| 101 | 5.01 | 5.43 | 6.39 |
| | *r* | 0.8962 | 0.9181 |
| | *s* | 0.51 | 0.51 |

[a] In the calculations we used the parameters given in Table 3. [b] The $\log P$ values were taken from [6] (in the other cases, from [1]).

the lowest rms deviation of the estimates from the experimental data. In this selection procedure, the structural sense of the terms of the correlation equation remains clear, which allows interpretation of the model from the viewpoint of what structural elements, and in what manner, affect the activity.

## REFERENCES


1. Hansch, C. and Anderson, S.M., *J. Med. Chem.,* 1967, vol. 10, no. 5, p. 745.
2. Hsieh, M.-M. and Dorsey, J.G., *Anal. Chem.,* 1995, vol. 67, no. 1, p. 48.
3. Cuenca-Benito, M., Sargado, S., Villanueva-Camanas, R.M., and Medina-Hernandez, M.J., *J. Chromatogr. A,* 1998, vol. 814, nos. 1–2, p. 121.
4. Jakab, A., Schulert, G., Prodan, M., and Forgacs, E., *J. Chromatogr. B,* 2002, vol. 770, nos. 1–2, p. 227.
5. Pyka, A., Kepczynska, E., and Bojarski, J., *Indian J. Chem., Sect. A,* 2003, vol. 42, no. 6, p. 1405.
6. Blakey, G.E., Nestorov, I.A., Arundel, Ph.A., Aarons, L.J., and Rowland, M., *J.Pharmacokinet. Biopharm.,* 1997, vol. 25, no. 3, p. 277.
7. Breyholz, H.-J., Schäfers, M., Wagner, S., Höltke, C., Faust, A., Rabeneck, H., Levkau, B., Scholer, O., and Kopka, K., *J. Med. Chem.,* 2005, vol. 48, no. 9, p. 3400.
8. Wells, M.J.M., Clark, C.R., and Patterson, R.M., *J. Chromatogr. Sci.,* 1981, vol. 19, no. 11, p.573.
9. Roca, R.M.S., March, F.G., Alvarez, J.G., and Domenech, R.G., *An. Quim.,* 1992, vol.88, no. 3, p. 382.
10. Stuper, A.J., Brügger, W.E., and Jurs, P.C., *Computer-Asisted Studies of Chemical Structure and Biological Function,* New York: Wiley, 1979. Translated under the title *Mashinnyi analiz svyazi khimicheskoi struktury i biologicheskoi aktivnosti,* Moscow: Mir, 1982, p. 165.
11. Golovanov, I.B. and Tsygankova, I.G., *Zh. Obshch. Khim.,* 1999, vol. 69, no. 8, p. 1275.
12. Golovanov, I.B. and Zhenodarova, S.M., *Zh. Obshch. Khim.,* 2003, vol. 73, no. 1, p. 90.
13. Golovanov, I.B., Zhenodarova, S.M., Ivanitskii, G.R., and Sedel'nikova, E.A., *Dokl. Ross. Akad. Nauk,* 2002, vol. 382, no. 5, p. 695.
14. Golovanov, I.B., Zhenodarova, S.M., Ivanitskii, G.R., and Khabarova, M.I., *Dokl. Ross. Akad. Nauk,* 2002, vol. 387, no. 1, p. 113.
15. Balabaev, N.K., Golovanov, I.B., Zhenodarova, S.M., Ivanitskii, G.R., Tsygankova,I.G., and Shnol', E.E., in *Gorizonty biofiziki* (Horizons of Biophysics), Pushchino, 2003, p. 206.
16. Golovanov, I.B. and Zhenodarova, S.M., *Zh. Obshch. Khim.,* 2005, vol. 75, no. 1, p. 82.
17. Rodionova, O.E., *Ross. Khim. Zh.,* 2006, vol. 50, no. 2, p. 128.
18. Draper, N.R. and Smith, H., *Applied Regression Analysis,* New York: Wiley, 1967.
19. Abdi, H., 2003. *http://www.utdallas.edu/~herve/.*
20. Golovanov, I.B., Ivanitskii, G.R., and Tsygankova, I.G., *Dokl. Ross. Akad. Nauk,* 1998, vol. 359, no. 2, p. 258.
21. Golovanov, I.B., Ivanitskii, G.R., and Tsygankova, I.G., *Dokl. Ross. Akad. Nauk,* 1998, vol. 359, no. 3, p. 409.
22. Golovanov, I.B., Ivanitskii, G.R., and Tsygankova, I.G., *Dokl. Ross. Akad. Nauk,* 1998, vol. 359, no. 6, p. 823.