ISSN 0006-2979, Biochemistry (Moscow), 2006, Vol. 71, No. 1, pp. 90-98. © Pleiades Publishing, Inc., 2006.
Original Russian Text © V. S. Sibirtsev, 2006, published in Biokhimiya, 2006, Vol. 71, No. 1, pp. 111-120.
Originally published in Biochemistry (Moscow) On-Line Papers in Press, as Manuscript BM05-097, December 4, 2005.

Analysis of Benzo[a]pyrene Deactivation Mechanisms in Rats

V. S. Sibirtsev

Mendeleyev Russian Research Institute for Metrology, Moskovskii pr. 19, 190005 St. Petersburg, Russia; fax: (7-812) 327-9776; E-mail: vs1969r@mail.ru

Received April 5, 2005
Revision received April 28, 2005

Abstract—The experimental data on the effects of a widespread carcinogen, benzo[a]pyrene (BP), on individual reactions of rats were treated using mathematical–statistical methods. The individual reactions were analyzed in dependence of doses and modes of administration (single or chronic). The analysis revealed a statistically significant correlation between life span and urinary content of (±)-*trans*-7,8-dihydroxy-7,8-dihydrobenzo[a]pyrene (7,8-BP) in rats treated with BP. The calculated regression equations revealed that the individual sensitivity to carcinogen in case of the BP single administration to rats is mainly determined by efficiency of excretion of the BP active forms out of the organism, whereas after chronic BP administration it is determined by mechanisms of enzymatic deactivation of BP.

DOI: 10.1134/S0006297906010147

Key words: benzo[a]pyrene, carcinogen deactivation, regression analysis, correlation analysis, interpolation, approximation, splines

Benzo[a]pyrene (BP) is one of polycyclic aromatic carbohydrates, widespread compounds playing a significant role in induction of various malignant tumors in man. Significant amounts of BP are formed during incomplete combustion of various organic materials. According to the report of IARC experts, products of BP metabolism are especially dangerous for man [1].

(±)-*Trans*-7,8-dihydroxy-7,8-dihydrobenzo[a]-pyrene (7,8-BP) is the main marker cancer metabolite. Subsequent metabolism of this compound yields diol-epoxide; the latter can interact with DNA and cause malignant transformation [2-4]. However, earlier studies revealed only a tendency (but not statistically significant correlation) between reduction of the latent period of tumor manifestation in rats and increase in 7,8-BP total urinary excretion after single dose administration of BP to animals (200 mg/kg in sunflower oil) [2-7]. The authors reported large individual variations of the parameters investigated suggesting heterogeneity of the animals; this factor together with standard errors of metabolic parameters assayed would account for lack of statistically significant changes.

Increasing the number of experimental animals would represent one possible approach for demonstration of statistically significant interrelationships between the measured parameters. However, this approach is rather time consuming and expensive. The major goal of this study was to evaluate significant individual sensitivity of experimental animals to different modes of BP administration and to find characteristic features of BP deactivation under its acute and chronic administration using a limited number of experimental animals. This required the use of a modified method for treatment of experimental data: abnormal value exclusion, correlation and regression analysis employing nonlinear dependences with coefficient numbers exceeding two, etc.

MATERIALS AND METHODS

The primary experimental data were obtained using male LIO rats [8] (Petrov Institute of Oncology) weighing 150 g at the beginning of experiments. Animals were treated with single doses of BP (100 or 200 mg/kg) or chronically (10 mg/kg BP × 10 times with 10-day interval between injections). BP was injected intraperitoneally as an aqueous–lipid emulsion.

It was shown earlier [2-7] that under these conditions of BP administration individual sensitivity to this compound expressed as the latent period from the beginning of BP administration to rats up to onset of first signs of neoplasm formation mainly depends on the urinary output of 7,8-BP. However, paired correlation coefficient values between these parameters were statistically

insignificant. Taking into consideration these results and also the reasonable suggestion that besides carcinogenic effect BP may exert general toxic effect on experimental animals, we have selected life span after the beginning of BP administration as the parameter characterizing individual sensitivity of rats to BP. This parameter was compared with total urinary output of 7,8-BP; urine was collected during the first 5 days after BP administration, and 7,8-BP was analyzed by the methods described in [2-7]. The time interval of 5 days is characterized by excretion of the major proportion of administered BP [2-7]. Rats were observed until their natural death.

The primary experimental data were treated as follows. In the first stage, we excluded "abnormal" values from consideration. Many researchers do not pay much attention to this stage, because they believe that the more experimental data they use the more accurate will be the result. But this is true not in all cases. Appearance of "abnormal" values may originate not only from the existence of some experimental errors during analyses of the parameters y and x but also from heterogeneity (especially of living organisms) of the research objects. It is also possible that y may also be influenced by factors other than x. However, this procedure has to be done very carefully because of possible cutoff of "critical" points, which are especially crucial as they carry qualitatively important information on altered behavior of the studied dependence at its various parts.

In the case of a large amount of primary experimental data, use of the following method would give the desired result. An initial set of results containing Q^* values of the parameters y and x, is treated using the following ratio:

$$L = (x_Q - x_1)/(y_Q - y_1).$$

If $L < 1$, change positions of the parameters y and x. Let $\Delta x = (x_Q - x_1)/Q$ (where $Q < Q^*$). For all Q^{**} values of x and y parameters of the initial set which fit to the Δx_i range ($i = 1 \div Q$) we find $x_i^* = \Sigma x_j/Q^{**}$ and $y_i^* = \Sigma y_j/Q^{**}$ ($j = 1 \div Q^{**}$); they form a final smoothed set containing Q but not Q^*, number of x_i^* and y_i^* parameters which are used in subsequent analysis.

However, our case required another approach: for each (y_j, x_j) point of the primary two-dimensional set (containing Q number of such points) b the nearest points are also determined. If the resultant ($b + 1$) value is consistent with the criterion:

$$|M_j - y_j| > t_\alpha \, [\Sigma(M_j - y_i)^2/(b^2 - b)]^{1/2}, \qquad (1)$$

where $M_j = \Sigma y_i/(b + 1)$, $i = 1 \div (b + 1)$, $i \neq j$, and t_α is Student's criterion for the significance level "α" and "$b - 1$" number of degrees of freedom, the point (y_j, x_j) is considered as abnormal and is excluded from subsequent calculation [9].

The second stage includes rough analysis of mutual dependence of y and x parameters. We used Spearman's coefficient of rank correlation as follows. Suppose that in the investigated set of results x_j and y_j values are positioned at s_i and z_i, respectively. If

$$(Q - 1)^{1/2} \cdot [1 - 6/(Q^3 - Q)] \cdot \Sigma(s_i - z_i)^2 > u_\alpha, \qquad (2)$$

where u_α is a table value of the Laplace function ($u_{0.1} = 0.2533$, $u_{0.05} = 0.125$, $u_{0.01} = 0.025$, $u_{0.001} = 0.0025$), the hypothesis on independence of y and x parameters is rejected at the significance level "α" [10].

If the analysis employing Spearman correlation coefficient (after exclusion of abnormal data) gives positive results, it is important to "visualize" this dependence. The simplest way is interpolation when the resultant curve passes through an ordered group of points of initial data [9-12].

In the case of linear interpolation, these points are connected with straight lines. For each two points (y_i, x_i) and (y_{i+1}, x_{i+1}) coefficients a_j ($j = 1 \div 2$) of the function $y = a_1 + a_2x$ can be found by solution of the system of two linear equations:

$$\begin{cases} a_1 + a_2 x_i = y_i \\ a_1 + a_2 x_{i+1} = y_{i+1}, \end{cases}$$

and finally this yields:

$$y = y_i + (x - x_i) \cdot (y_{i+1} - y_i)/(x_{i+1} - x_i). \qquad (3)$$

However, this method has limited applicability if the initial data are variable and heterogeneous. In this case, it is possible to increase the degree of the interpolating polynomial. In the case of cubic interpolation the coefficients a_j ($j = 1 \div 4$) of the function

$$y = a_1 + a_2x + a_3x^2 + a_4x^3$$

are found for each of two points (y_i, x_i) and (y_{i+1}, x_{i+1}) (which do not represent the end-point of the whole data set) by solving the system of four linear (versus a_j) equations:

$$\begin{cases} a_1 + a_2 x_{i-1} + a_3 (x_{i-1})^2 + a_4 (x_{i-1})^3 = y_{i-1} \\ a_1 + a_2 x_i + a_3 x_i^2 + a_4 x_i^3 = y_i \\ a_1 + a_2 x_{i+1} + a_3 (x_{i+1})^2 + a_4 (x_{i+1})^3 = y_{i+1} \\ a_1 + a_2 x_{i+2} + a_3 (x_{i+2})^2 + a_4 (x_{i+2})^3 = y_{i+2}. \end{cases}$$

By solving the system of linear (versus a_j) equations $a_1 + \Sigma a_j f_j(x_i) = f_0(y_i)$ (the most widespread equations are: $f(s) = s$, $1/s$, s^h, h^s, $log_h s$, $1/(h + s)$, $h_1/(h_2 + s)$, etc., where

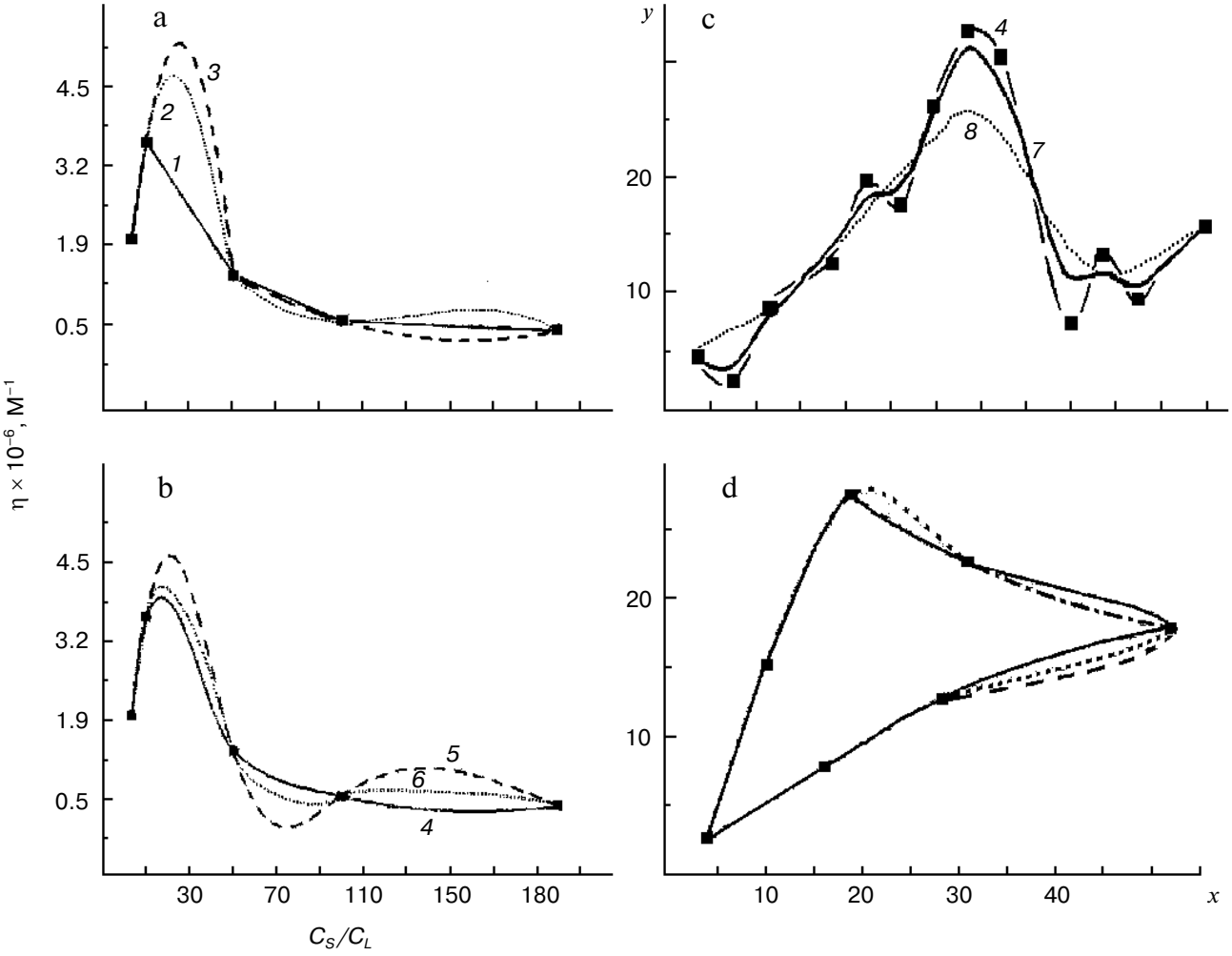

Fig. 1. Examples illustrating the use of various types of interpolation and smoothing for graphic presentation of the dependence between two parameters based on a limited number of observations (designated by dots). a, b) Initial data have been taken from the paper by Sibirtsev et al. [13] for the system "6-(2-imidazolin-2-yl-2-[4-(2-imidazolin-2-yl)phenyl]indole + calf thymus DNA" in an aqueous buffer containing 0.01 M NaCl + 0.01 M Na_2EDTA (sodium ethylene diamine tetraacetate) + 0.01 M Tris (2-amino-2-(hydroxymethyl)-1,3-propanediol), pH 7.4. The abscissa shows ratios of molar concentrations of DNA and a dye in the system (C_S/C_L), the ordinate shows values of fluorescent sensitivity coefficient reflecting augmentation of dye fluorescence during increase in DNA in the system per 1 M (η). c) Data were modeled by the author. d) An example of generation of closed dependence by means of interpolation of experimental data by locally set square splines with selection of various "set-points" and directions of movement along other points versus the selected set-point (clockwise or anticlockwise, respectively). Curves *1-3* show the result of interpolation of the experimental dataset using "normal" locally fitted polynomials of the first (see formula (3)), the second, and the third degrees, respectively. Curve *4* shows the result of interpolation of experimental data locally set by square splines (see formula (4)). Curves *5* and *6* show the result of interpolation of experimental data by "normal" and "tensed" globally fitted cubic spline (see formulas (5) and (7)). Curves *7* and *8* show smoothing of experimental data by various methods (see formulas (8), (9), etc.).

h_1 and h_2 = const) it is possible to find coefficient values for any other interpolation function type: $f_0(y) = a_1 + \Sigma a_j f_j(x)$. However, this approach cannot provide the desired result in all cases (Fig. 1). In such cases the interpolation splines, providing a continuum for the data $y(x_i) = y_i$, but also providing continuum for the first, the second order derivatives, etc. (in dependence of spline degree) are applicable [11, 14-16]. In this case as well as in the case of polynomial interpolation, the resultant function can be set either "globally" (using the whole dataset) or locally (taking into consideration only data surrounding a particular point).

Finding of a_j coefficients of local spline of

$$f_0(y) = z = a_0 + a_1 f_1(x) + a_2 f_2(x) + a_3 f_3(x),$$

providing a continuum of zero and the first order derivatives of the interpolation function within the range

between (y_i, x_i) and (y_{i+1}, x_{i+1}) $(i = 1 \div Q)$ points of the dataset requires solution of the following linear equations versus these coefficients:

$$\begin{cases} a_0 + a_1 f_1(x_i) + a_2 f_2(x_i) + a_3 f_3(x_i) = z_i \\ a_0 + a_1 f_1(x_{i+1}) + a_2 f_2(x_{i+1}) + a_3 f_3(x_{i+1}) = z_{i+1} \\ a_0 + a_1 f_1'(x_i) + a_2 f_2'(x_i) + a_3 f_3'(x_i) = d_i \\ a_0 + a_1 f_1'(x_{i+1}) + a_2 f_2'(x_{i+1}) + a_3 f_3'(x_{i+1}) = d_{i+1}, \end{cases}$$

where $f_j'(x)$ $(j = 1 \div 3)$ are the first order derivatives of $f_j(x)$, and d_j $(j = i \div (i+1))$ are the first order derivatives of z_j determined, for example, by formulas of differential ratios [9-11] by three points:

$d_i = (z_{i+1} - z_i)/(x_{i+1} - x_i)$, if $i = 1$ or $(x_{i+1} - x_i) < 2(x_i - x_{i-1})$,

$d_i = (z_i - z_{i-1})/(x_i - x_{i-1})$, if $i = Q$ or $(x_{i+1} - x_i) > 2(x_i - x_{i-1})$,

$d_i = (z_{i+1} - z_{i-1})/(x_{i+1} - x_{i-1})$ in the other cases.

For a particular form of the interpolation function:

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

all these constructions yield the final expression [11]:

$$y(x) = [d_i (x_{i+1} - x)^2 (x - x_i) + d_{i+1} (x - x_i)^2 (x - x_{i+1})]/$$
$$/(x_{i+1} - x_i)^2 + [y_i (x_{i+1} - x)^2 (2x - 3x_i + x_{i+1}) +$$
$$+ y_{i+1} (x - x_i)^2 (3x_{i+1} - x_i - 2x)]/(x_{i+1} - x_i)^3. \quad (4)$$

The local splines for each point (y_i, x_i) can also be set by V neighboring points of the whole dataset rather than by one point. In this case, the interpolating polynomial spline has the following form:

$$f_0(y) = z = a_1 + \Sigma a_j f_j(x) \text{ (usually } y = a_1 + \Sigma a_j x^{j-1}),$$

where $j = 2 \div 2V$ for splines with continuous only zero and the first order derivatives, or $j = 2 \div 3V$ for splines with continuous zero, first, and second order derivatives; and the system of equations for splines with continuous zero and first order derivatives solved versus a_j has the following form:

$$\begin{cases} a_1 + \Sigma a_j f_j(x_{i-V}) = z_{i-V} \\ \dots \\ a_1 + \Sigma a_j f_j(x_{i+V}) = z_{i+V} \\ a_1 + \Sigma a_j f_j'(x_{i-V}) = d_{i-V} \\ \dots \\ a_1 + \Sigma a_j f_j'(x_{i+V}) = d_{i+V}, \end{cases}$$

usually

$$\begin{cases} a_1 + \Sigma a_j (x_{i-V})^{j-1} = z_{i-V} \\ \dots \\ a_1 + \Sigma a_j (x_{i+V})^{j-1} = z_{i+V} \\ a_1 + \Sigma a_j (j-1)(x_{i-V})^{j-2} = d_{i-V} \\ \dots \\ a_1 + \Sigma a_j (j-1)(x_{i+V})^{j-2} = d_{i+V}, \end{cases}$$

and for splines with continuous zero, first, and second order derivatives the system of equations has the following form:

$$\begin{cases} a_1 + \Sigma a_j f_j(x_{i-V}) = z_{i-V} \\ \dots \\ a_1 + \Sigma a_j f_j(x_{i+V}) = z_{i+V} \\ a_1 + \Sigma a_j f_j'(x_{i-V}) = d_{i-V} \\ \dots \\ a_1 + \Sigma a_j f_j'(x_{i+V}) = d_{i+V} \\ a_1 + \Sigma a_j f_j''(x_{i-V}) = d_{i-V}^* \\ \dots \\ a_1 + \Sigma a_j f_j''(x_{i+V}) = d_{i+V}^*, \end{cases}$$

usually

$$\begin{cases} a_1 + \Sigma a_j (x_{i-V})^{j-1} = z_{i-V} \\ \dots \\ a_1 + \Sigma a_j (x_{i+V})^{j-1} = z_{i+V} \\ a_1 + \Sigma a_j (j-1)(x_{i-V})^{j-2} = d_{i-V} \\ \dots \\ a_1 + \Sigma a_j (j-1)(x_{i+V})^{j-2} = d_{i+V} \\ a_1 + \Sigma a_j (j-1)(j-2)(x_{i-V})^{j-3} = d_{i-V}^* \\ \dots \\ a_1 + \Sigma a_j (j-1)(j-2)(x_{i+V})^{j-3} = d_{i+V}^*, \end{cases}$$

where $f_j''(x)$ $(j = 1 \div 3)$ are the second derivatives from $f_j(x)$ and d_j^* are the second derivatives from z_i determined for example by formulas of differential ratios [9-11] by three points:

$d_i^* = (d_{i+1} - d_i)/(x_{i+1} - x_i)$, if $i = 1$ or $(x_{i+1} - x_i) < 2(x_i - x_{i-1})$,

$d_i^* = (d_i - d_{i-1})/(x_i - x_{i-1})$, if $i = Q$ or $(x_{i+1} - x_i) > 2(x_i - x_{i-1})$,

$d_i^* = (d_{i+1} - d_{i-1})/(x_{i+1} - x_{i-1})$ in all other cases.

The "global" spline of the third order providing continuum for zero, the first, the second, and the third order derivatives of the interpolation function can be set for each $x_i \div x_{i+1}$ range, for example, in the form [11]:

$$y(x) = [y_i/(x_{i+1} - x_i) - a_i(x_{i+1} - x_i)/6]\cdot(x_{i+1} - x) +$$

$$+ [y_{i+1}/(x_{i+1} - x_i) - a_{i+1}(x_{i+1} - x_i)/6]\cdot(x - x_i) +$$

$$+ [a_i(x_{i+1} - x)^3 + a_{i+1}(x - x_i)^3 (x - x_{i+1})]/[6(x_{i+1} - x_i)], \quad (5)$$

where $a_1 = a_Q = 0$, and a_j can be found by solving the following system of linear equations:

$$a_i(x_{i+1} - x_i) + 2a_{i+1}(x_{i+2} - x_i) + a_{i+2}(x_{i+2} - x_{i+1}) =$$

$$= (y_{i+2} - y_{i+1})/[6(x_{i+2} - x_{i+1})] - (y_{i+1} - y_i)/[6(x_{i+1} - x_i)].$$

Asymptotic behavior of the interpolation function is usually set [11] in the following form:

$$y(x) = y_1 + d_1(x - x_1) \text{ at } x < x_1;$$

$$\text{or } y(x) = y_Q + d_Q(x - x_Q) \text{ at } x > x_Q. \quad (6)$$

For construction a closed or self-crossed curve, the interpolation function is set by the "local" mode; the latter includes analysis of each two (or more if necessary) adjacent points of the dataset by the following ratio:

$$L = (x_{i+1} - x_i)/(y_{i+1} - y_i).$$

If $L < 1$, x versus y is interpolated and in the above formulas y and x just change their positions; if $L > 1$, y versus x is interpolated.

In the case of rather distant positioning of experimental points and in some other cases, which require exclusion (or at least reduction) of abnormal behavior (e.g. extremum points characterized by change of sign of the first order derivative, bent points characterized by sign changes of the second order derivative), it may be useful to employ so-called tension bracing functions [16]. The resultant y value at the given x value is calculated using the following equation:

$$y(x) = y_1(x) + L^*[y_2(x) - y_1(x)], \quad (7)$$

where $y_1(x)$ is determined by the initial spline (see formulas (4) and (5), for example) or any other interpolation function; $y_2(x)$ is determined by the "rod" linear (see formula (3)) or any other interpolation function; L^* is a "tension" coefficient and its values ranging from 0 to 1 can be arbitrary selected by researchers for particular range $x_i \div x_{i+1}$ of the analyzed dataset.

If the initial data are not perfectly reliable the analyzed dataset may be subjected to pilot smoothing; after this treatment the modified parameters x and y can be used for construction of the resultant interpolation factor. This can be achieved using several approaches [14]. For example, for each i-point of the initial dataset the interpolation factor may be constructed by V values of the nearest points (but ignoring the i-point itself) and then the resultant x_i value can be used for calculation of a new y_i^* value. In the case of local linear interpolation the y_i^* value can be calculated as follows:

$$y_i^* = y_{i-1} + (x_i - x_{i-1})\cdot(y_{i+1} - y_{i-1})/(x_{i+1} - x_{i-1}) \quad (8)$$

(compare with Eq. (3)).

The other way includes determination of modified y_i^{**} value at a given x_i and y_i values as:

$$y_i^{**} = y_i(1 - L^*) + y_i^* L^*, \quad (9)$$

where y_i^* is the smoothed y_i value determined as described above, a L^* is the correction coefficient ranging from 0 to 1.

The other mode of data smoothing can consist in construction of the interpolation factor using any of the above described methods and the initial dataset. Dividing the initial dataset into Q^* ranges possessing identical length of x-parameter (Δx) it is possible to calculate y_{ij}^{**} values for each of these Q^{**} ranges using x_{ij}^{**} positioned at identical length from each other. The resultant data are then averaged by the chosen ranges in the following form:

$$x_i^* = \Sigma x_{ij}^{**}/Q^{**}$$

$$\text{and } y_i^* = \Sigma y_{ij}^{**}/Q^{**} \ (i = 1 \div Q^*, j = 1 \div Q^{**}).$$

Finally, the fourth way of data smoothing consists in construction of a bracing tension interpolation factor (see formula (7)), where $y_1(x)$ is constructed in the usual way using the initial dataset. Equation (8) or similar equations for 4, 6, etc. points nearest to the smoothing one can be used for $y_2(x)$ (the rod function). In the nodes of the final interpolation factor y_i^{**} will be determined by formula (9). It is also possible to use approximation (even with low reliability) of initial data (see below) rather than interpolation.

If results of smoothing are not satisfactory, it is advisable to repeat this procedure with initial and new (smoothed) data using the same or previously described method with the same or other smoothing parameters. These include V or Δx determining the smoothing range; the correcting coefficient L^*; interpolation factors of the initial datasets and factors setting the rod function $y_2(x)$, etc. Certain care should be taken for selection of these factors.

Finally at the fourth stage after graphic presentation of the investigated dependence (Fig. 1) it is possible to try to determine this dependence in the form of a function with limited (as minimal as possible) number of coefficients; it is close to each point of the analyzed dataset (but not necessarily to pass through each point). This is an approximation procedure. If a law determining the form

of required dependence is unknown (otherwise interpolation is not required and in this case it is recommended just to find coefficients of requested dependence) approximation should be started from the simplest law:

$$f_0(y) = a_0 + a_1 f_1(x), \quad (10)$$

where $f_0(y)$ and $f_1(x)$ are chosen to give the graphic presentation of the dependence (10) by the form and behavior resembling that obtained in the previous stage. (The most widespread of $f_0(y)$ and $f_1(x)$: $f(s) = s$, $1/s$, s^h, h^s, $log_h s$, $1/(h + s)$, $h_1/(h_2 + s)$, etc., where h_1 and $h_2 = \text{const}$).

Replacement of variables $z = f_0(y)$, $g = f_1(x)$ transforms this dependence into the following form:

$$z = a_0 + a_1 g. \quad (11)$$

The correlation coefficient for this can be calculated using the standard formula:

$$r = \frac{\Sigma(w_i^2 g_i z_i) - \Sigma(w_i g_i)\Sigma(w_i z_i)/Q}{([\Sigma(w_i g_i)^2 - (\Sigma w_i g_i)^2/Q]\,[\Sigma(w_i z_i)^2 - (\Sigma w_i z_i)^2/Q])^{1/2}}, \quad (12)$$

and the regression coefficients can be calculated by formulas [11]:

$$a_1 = \frac{\Sigma(w_i g_i)\Sigma(w_i z_i) - Q\Sigma(w_i^2 g_i z_i)}{\Sigma(w_i g_i)^2 - Q\Sigma(w_i g_i)^2},$$

$$a_0 = 1/Q\,[\Sigma(w_i z_i) - a_1\Sigma(w_i g_i)], \quad (13)$$

where w_i are weight coefficients. Their values ranging from 0 to 1 reflect reliability of determination of i-point in the analyzed dataset. These values can be determined by the researchers themselves (e.g. as $w_i = 1$ at identical significance of all the data analyzed) or using methods given elsewhere (e.g. in [17]). Linear correlation between z and g is considered validly established with the probability $P = 100(1 - \alpha)$, if $|r| > r_\alpha$, where critical value of the correlation coefficient is determined as:

$$r_\alpha = t_\alpha/[t_\alpha^2 + Q - 2]^{1/2} \quad (14)$$

(here t_α is the table value of Student's criterion for the significance level "α" and "$Q - 2$" number of the degrees of freedom).

Extension of this approach to the dependence:

$$f_0(y) = a_0 + a_1 f_1(x) + \ldots + a_k f_k(x), \quad (15)$$

describing k interrelated processes in the body can be achieved by replacing variables:

$$z = f_0(y) \text{ and } g = f_1(x) + (a_2/a_1)f_2(x) + \ldots + (a_k/a_1)f_k(x). \quad (16)$$

After that the correlation coefficient (r) can be calculated using formula (12).

Calculation of regression coefficients $a_0 \div a_k$ of the dependence (15) requires solution of the system of linear equations:

$$\begin{cases} Qa_0\Sigma w_i + a_1\Sigma[w_i f_1(x_i)] + \ldots + a_k\Sigma[w_i f_k(x_i)] = \\ = \Sigma[w_i f_0(y_i)] \\ a_0\Sigma[w_i f_1(x_i)] + a_1\Sigma[w_i f_1(x_i)]^2 + \ldots + \\ + a_k\Sigma[w_i^2 f_k(x_i) f_1(x_i)] = \Sigma[w_i^2 f_0(y_i) f_1(x_i)] \\ \ldots \\ a_0\Sigma[w_i f_k(x_i)] + a_1\Sigma[w_i^2 f_1(x_i) f_k(x_i)] + \ldots + \\ + a_k\Sigma[w_i f_k(x_i)]^2 = \Sigma[w_i^2 f_0(y_i) f_k(x_i)] \end{cases}$$

versus $a_0 \div a_k$ using the least squares method [18].

The form of functions $f_0(y)$, $f_1(x) \div f_k(x)$ and their numbers are chosen by researchers based on theoretical background of the behavior of the described process, maximization requirements for calculation of r/r_α and K_{ad} (regression equation adequacy factor) and also minimization of relative approximation error (ε). Here r_α is determined by formula (14) for dependences (10) and (11); for dependence (15) it is determined by the following formula:

$$r_\alpha = [1/(1 + (Q - k + 1)/[kF_\alpha])]^{1/2} \quad (17)$$

(where F_α is the table value of the Fisher criterion for the significance level "α" and number of the degree of freedom "k" and "$Q - k + 1$", respectively).

K_{ad} and ε are determined by the formulas [19]:

$$K_{ad} = (Q - k)\cdot\Sigma(y_i^T)^2/[kF_\alpha\Sigma(y_i^T - y_i^E)^2], \quad (18)$$

$$\varepsilon = (100/Q)\cdot\Sigma|(y_i^T - y_i^E)/y_i^E| \quad (19)$$

(where y_i^T and y_i^E are theoretical (calculated by formula (15)) and experimental values of parameter y, respectively).

Another way of analysis of dependence (15) consists in its consideration as a function of more than x variable numbers. In Eq. (15) (and related equations), "$f_0(y)$, $f_1(x)$ ... $f_k(x)$" should be replaced by "$f_0(x_1)$, $f_1(x_2 ... x_M)$ ... $f_k(x_2 ... x_M)$" (here y is considered as x_1).

Paired (r_{jp}) and general (R_0) correlation coefficients are calculated as in the case of two-dimensional analysis of the constituent function but using the following replacements: for (r_{jp}) calculation only "x_{ji}" and "x_{pi}" replace "g_i" and "z_i" in the formula (12); for R_0 calculation "$f_0(x_1)$, $f_1(x_2 ... x_M)$ ... $f_k(x_2 ... x_M)$" replace "$f_0(y)$, $f_1(x)$ ... $f_k(x)$" in formula (16). During consideration partial and multiple correlation coefficients appear [19]. The partial

correlation coefficients r_{jp}* characterize a degree of linear relationship between x_j and x_p parameters (excluding linear influences of other x_l ($l = 1 \div M$, $l \neq j$, $l \neq p$) parameters included into consideration). The multiple correlation coefficients R_j characterize a degree of linear relationship between x_j parameter and other x_l ($l = 1 \div M$, $l \neq j$) parameters included into consideration. Partial correlation coefficients are calculated by the formula [10]:

$$r_{jp}^* = -\Lambda_{jp}/(\Lambda_{jj}\Lambda_{pp})^{1/2}, \quad (20)$$

where $[\Lambda_{jp}] = [\Lambda_{jp}^*]^{-1}$ is the matrix, inversed to the second order central moment matrix:

$$\Lambda_{jp}^* = \Sigma(x_{ji} - \Sigma x_{ji}/Q)/(x_{pi} - \Sigma x_{pi}/Q),\ i = 1 \div Q, \quad (21)$$

and Q are total points in the dataset. And linear correlation between x_j and x_p parameters at exception of linear influence on them of other x_l ($l = 1 \div M$, $l \neq j$, $l \neq p$) parameters is considered validly established with the probability $P = 100(1 - \alpha)$, if

$$|r_{jp}^*(Q - M - 4)/[1 - (r_{jp}^*)^2]^{1/2}| > t_\alpha \quad (22)$$

(where t_α is the table value of Student's criterion at the significance level "α" and "$Q - M - 4$" number of degrees of freedom, and M is the "dimension" of the dataset) [19].

The multiple correlation coefficients are calculated by formula [10]:

$$R_j = [1 - 1/(\Lambda_{jj}\Lambda_{jp}^*)]. \quad (23)$$

And linear correlation between x_j and other x_l ($l = 1 \div M$, $l \neq j$) parameters is considered validly established with the probability $P = 100(1 - \alpha)$, if

$$|R_j| > [1/(1 + (Q - M + 1)/[kF_\alpha])]^{1/2} \quad (24)$$

(where F_α is the table value of the Fisher criterion for the "α" significance level and "M" and "$Q - M + 1$" number of degrees of freedom).

RESULTS AND DISCUSSION

After exclusion of abnormal values using the methods described above the statistically significant relationship ($\alpha < 0.05$) was found between the increase in individual life span (τ) of rats treated with a single dose of BP (100 or 200 mg/kg) and the individual increase in total urinary excretion of the marker metabolite 7,8-BP during the first five days after BP administration (C_{BP}). This relationship was either linear (in general for these groups) or exponential (for animals in which BP administration caused tumors of internal organs) (see Fig. 2). Lack of significant changes between these curves suggests equal contribution of carcinogenic and toxic effects of 7,8-BP into shortening of life span of rats. The increase in τ accompanied by the corresponding increase in C_{BP} suggests that life span of animals treated with a single dose of BP is directly linked to efficacy of excretion of active metabolite 7,8-BP formed in the body. However, the two-fold increase in the single dose of BP from 100 to 200 mg/kg at the urinary excretion of the same amount of its metabolite, 7,8-BP, was accompanied by four-fold reduction in life span of rats ($\tau = 12C_{BP}$ and $\tau = 3C_{BP}$ for administration of 100 and 200 mg/kg BP, respectively; see Fig. 2). Consequently, the amount of 7,8-BP formation in the body is characterized by nonlinear dependence on the dose of BP administered.

In the case of chronic administration of BP to rats, their reactions to this xenobiotic changed. Even using the

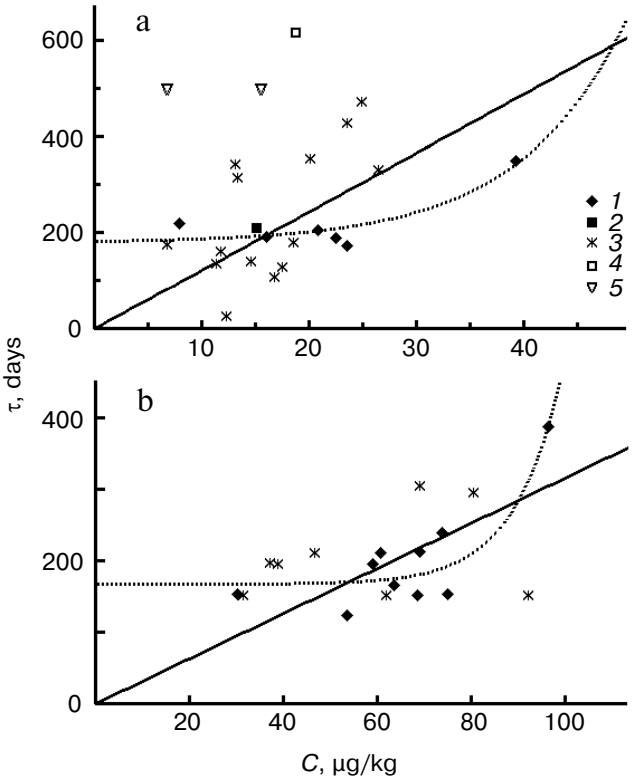

Fig. 2. Dependence of individual life span (τ) of rats on urinary excretion of 7,8-BP (C) during the first five days after a single administration of BP in the dose 100 (a) and 200 (b) mg/kg. Points *1-3* show experimental data for animals treated with BP and characterized by: *1*) internal organ tumor formation; *2*) subcutaneous tumor formation; *3*) lack of tumors. Points *4* and *5* show data cutoff for the last two groups according to formula (1). Solid curves designate calculated regression equations for the whole group of animals ($\tau = 12.2C$, $r = 0.510$, $\alpha < 0.05$ (a); $\tau = 3.15C$, $r = 0.484$, $\alpha < 0.05$ (b), where r is the correlation coefficient and α is significance level). Dotted curves designate calculated regression equations for animals characterized by tumor formation after BP administration ($\tau = 178 + 3.23 \cdot e^{C/10}$, $r = 0.917$, $\alpha < 0.01$ (a); $\tau = 167 + 0.014 \cdot e^{C/10}$, $r = 0.893$, $\alpha < 0.01$ (b)).

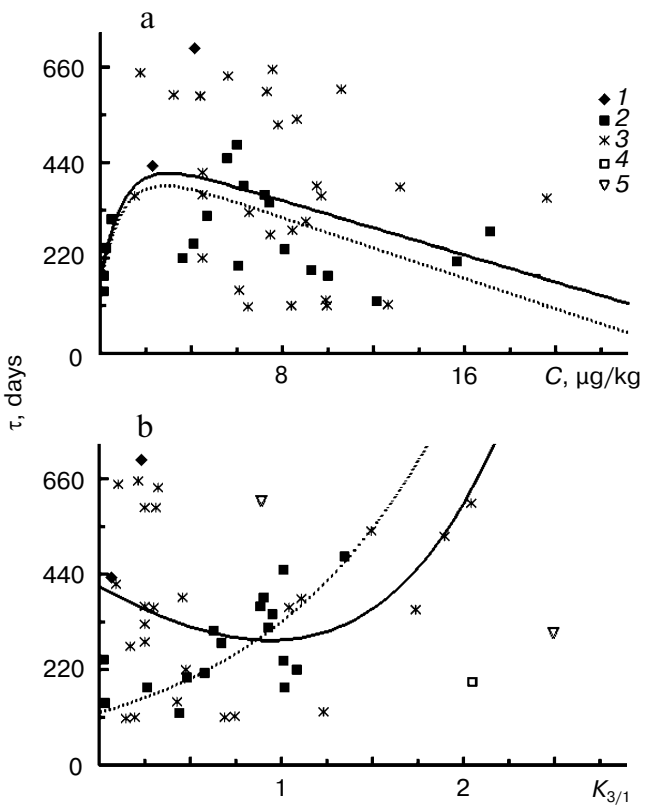

Fig. 3. Dependence of individual life span (τ) of rats either on urinary excretion of 7,8-BP (*C*) during the first five days after the third BP injection (a) or on the ratio of total urinary content of 7,8-BP during the first five days after the third administration of BP to the 7,8-BP content during the first five days after the first administration of BP ($K_{3/1}$) (b) (on chronic administration of BP to rats that received 10 injections of BP in the dose 10 mg/kg with 10-day intervals). Solid curves designate calculated regression equations for the whole group of animals ($\tau = 477 - 15.6C - 309e^{-C}$, $r = 0.372$, $\alpha < 0.05$ (a); $\tau = 265 - 377K + 147e^{K}$, $r = 0.392$, $\alpha < 0.05$ (b)). Dotted curves designate calculated regression equations for animals characterized by subcutaneous tumor formation after BP administration ($\tau = 454 - 17.5C - 296e^{-C}$, $r = 0.547$, $\alpha < 0.05$ (a); $\tau = 121e^{K}$, $r = 0.663$, $\alpha < 0.01$ (b)). All other designations for the points *1-5* are the same as in Fig. 2.

Spearman range correlation criterion (see formula (2)) we failed to find any dependence of τ of animals that received 10 injections of BP in the dose 10 mg/kg on the value of partial urinary excretion of 7,8-BP within five days after the first and the second injections ($C_{BP,1}$ and $C_{BP,2}$, respectively). Such dependence was detected only after the third injection of BP. However, even in this case the increase in τ with the increase in $C_{BP,3}$ consistent with experiments on a single dose administration of BP was found only for $C_{BP,3} < 3$ μg/ml; at $C_{BP,3} > 3$ μg/ml, τ changed in the opposite direction with the increase in $C_{BP,3}$ (Fig. 3a).

These differences can be possibly attributed to the fact that primary reaction on the chronic administration of BP does not necessarily reflect individual adaptation possibilities of the body. After the second injection (10 days after the beginning of chronic BP administration) the adaptation mechanisms begin to react to the presence of BP in the body, and only after the third injection (20 days after the beginning of chronic BP administration to rats) the adaptation mechanisms include not only excretion of BP from the body (dominating at $C_{BP,3} < 3$ μg/ml) but also inducible mechanisms of its enzymatic deactivation. These inducible mechanisms dominate and in this case, as in the case of single dose administration of BP, the urinary excretion of 7,8-BP is directly proportional to the amount of BP metabolite in the body, which determines life span of experimental animals. However, the amount of the major carcinogenic metabolite depends not only on the amount of BP administered but also on the activity of BP-metabolizing enzymes. It should be noted that as in the case of single dose administration of BP, there were insignificant differences in the relationship between 7,8-BP excretion and life span in rats with and without malignant tumors formed (Fig. 3).

The ratio of partial urinary excretion of 7,8-BP during the first five days after the first and the third injections ($K_{3/1} = C_{BP,3}/C_{BP,1}$). Figure 3b shows the dependence of individual life span (τ) of rats chronically treated with BP on this parameter. For the group of animals studied τ changed with the increase in $K_{3/1}$ in a non-monotonous manner and decreased to $K_{3/1} = 1$ and then increased exponentially. However, in rats with subcutaneous tumors appearing after BP treatment τ increased with $K_{3/1}$ increase in exponential manner over the whole range. The latter group of animals was characterized by higher coefficient of correlation between τ and $K_{3/1}$ ($r = 0.663$, $\alpha < 0.01$) than that of correlation between τ and $C_{BP,3}$ ($r = 0.547$, $\alpha < 0.05$). It should be noted that description of the dependence of τ on $C_{BP,3}$ requires calculation of three parameters ($\tau = a_0 + a_1C + a_2e^{-C}$), whereas the dependence of τ on parameter $K_{3/1}$ requires calculation of only one ($\tau = a_1e^{K}$). This significantly simplifies calculation and additionally increases their reliability. These results might be extrapolated to man after evident correction on difference of BP metabolism in rat and man.

REFERENCES

1. *IARC Monographs on the Evaluation of Carcinogenic Risk to Humans* (1987) Suppl. 7, IARC, Lyon, France.
2. Likhachev, A. J., Zabezhinskii, M. A., and Dikun, P. P. (1992) *Vopr. Onkol.*, **38**, 950-956.
3. Likhachev, A. J., Beniashvili, D. Sh., and Bykov, V. Yu. (1992) *Environ. Health Perspect.*, **98**, 211-214.
4. Likhachev, A. J., Zabezhinskii, M. A., and Bykov, V. Yu. (1994) *Cancer Lett.*, **78**, 163-170.
5. Savochkina, I. V., Dikun, P. P., Zabezhinskii, M. A., Yamshchanov, V. A., and Likhachev, A. J. (1990) *Vopr. Onkol.*, **36**, 1221-1226.
6. Zabezhinskii, M. A., Likhachev, A. J., and Bykov, V. Ju. (1994) *Eksp. Onkol.*, **16**, 431-436.

7. Tyndyk, M. L. (1996) *Study of Forecast in Opportunity for Individual Sensitivity to Cancer Action of Benzo(a)pyrene on Excretion Its Metabolites*: Candidate's dissertation [in Russian], Petrov Oncological Research Institute, St. Petersburg.
8. Anisimov, V. N., Pliss, G. B., Iogannsen, M. G., Popovich, I. G., and Romanov, K. P. (1989) *J. Exp. Clin. Cancer Res.*, **8**, 324-330.
9. Kalitkin, N. N. (1978) *Numerical Methods* [in Russian], Fizmatgiz, Nauka, Moscow.
10. Korn, G., and Korn, T. (1968) *Mathematical Handbook for Scientists and Engineers. Definitions, Theorems and Formulas for Reference and Review*, McGraw-Hill Book Company.
11. Diakonov, V. P. (1987) *Reference Book on Algorithms and Programs in "Basic" Language for Personal Computer* [in Russian], Fizmatgiz, Nauka, Moscow.
12. Grinchishin, Ya. T., Efimov, V. I., and Lomakovich, A. N. (1988) *Algorithms and Programs on a "Basic"* [in Russian], Prosveshchenie, Moscow.
13. Sibirtsev, V. S., Garabadgiu, A. V., and Ivanov, S. D. (1994) *Bioorg. Khim.*, **20**, 650-668.
14. Xardle, V. (1990) *Applied Nonparametric Regression*, Cambridge University Press, New York.
15. Shikin, E. V., Boreskov, A. V., and Zaitsev, A. A. (1993) *Basis of the Computer Graphic* [in Russian], Dialog-MIFI, Moscow.
16. Shishkin, E. V., and Plis, L. I. (1996) *Curve and Surface on a Computer Screen. Manual on Splines for the Users* [in Russian], Dialog-MIFI, Moscow.
17. Johnson, K., and Jeffi, V. (1983) *Numerical Methods in Chemistry*, Cambridge University Press, New York.
18. Kafarov, V. V. (1972) *Programming and Computing Methods in Chemistry and Chemical Technology* [in Russian], Nauka, Moscow.
19. Dubrov, A. N., Mhatarjan, A. S., Troshin, L. I., and Maslechenko, I. V. (1991) *Mathematical-Statistical Analysis on Programming Microcalculators* [in Russian], Finansy i Statistika, Moscow.