

## **ON THE APPLICATION OF THE LONG-CHAIN APPROXIMATION FOR ESTIMATING BASICITY AND FIRST ELECTRONIC TRANSITION WAVELENGTH IN LARGE ARRAYS OF POLYMETHINE DYES**

**M. Dekhtyar**

Institute of Organic Chemistry, National Academy of Sciences of Ukraine, Murmanskaya 5,  
02094, Kiev-94, Ukraine

(Received July 1, 2005)

For a large array of polymethine dyes (about 100 carbocyanines with end nuclei belonging to 6 structural series), effective topological parameters of the long-chain approximation are calculated (electron donor ability  $\Phi_0$  and effective length  $L$ ) and correlated with experimental chemical and spectral dye characteristics (basicity thermodynamically determined by the negative logarithm of the conjugated acid dissociation constant  $pK_a$  and first electronic transition wavelength  $\lambda_{\text{exp}}$ ). A stable correlation is established between the series  $\Phi_0$  and  $pK_a$  ( $R = 0.96$ ). The  $\Phi_0$ -dependence of  $pK_a$  is found which has a general linear character on the whole basicity interval for the selected cyanine array and enables the prediction of the  $pK_a$  values for new dyes based on the  $\Phi_0$  values of their end groups as well as the solution of the inverse problem. A moderate coefficient of the correlation between  $\lambda_{\text{th}}$  calculated by  $L$  and  $\lambda_{\text{exp}}$  (0.84) suggests that the effective length of end nuclei is an appropriate quantity rather for finding trends in dye absorption maxima than for quantitative prognostication. Two ways of approximating the experimental values of dye absorption wavelengths are considered which involve homographic and linear functions of  $L$ . The improvement of  $\lambda_{\text{exp}}$  estimation using the corrected effective length values is discussed.

### **I. Introduction**

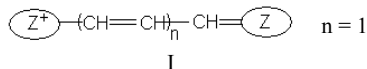
To purposefully design polymethine dyes (PMDs), it is necessary to have an algorithm for estimating the significant properties of new, sometimes hypothetical, heterocyclic nuclei involved as PMD end groups. This task implies determination of some theoretical (quantum chemical or empirical) parameters for a large number of both known and novel heterocyclic nuclei and establishment of quantitative relationships between such parameters and experimental PMD characteristics. A wealth of experience has been accumulated to date in the application of correlation analysis to studies on structure-property relationships of chemical compounds and it is

efficiently used in molecular design of materials with prescribed properties (see, e.g., Refs. 1, 2). In a particular case of PMDs, correlation analysis tools were employed to establish structural dependences of dye physicochemical and photographic behaviour [3]. Of most practical importance is the prediction of the first electronic transition wavelength,  $\lambda$ , and the  $pK_a$  value, since these characteristics represent a pivot of spectral and chemical nature of PMDs and hence underlie their technical uses.

Special interest is attached to the simplest prognosis methods which, unlike standard quantum chemical techniques, reveal not only numerical but also analytical relationships between integral molecular parameters and constitution of dyes. A method of this kind used here is an analytical approximation of the Hueckel method referred to as the long-chain approximation (LCA); its fundamentals were elaborated three decades ago, with the aim to describe quasi-one-dimensional conjugated systems embracing a wide diversity of polymethine compounds including PMDs [4-6]. LCA was extensively employed and proved quite efficient in estimating  $\pi$ -electronic properties of conjugated compounds, especially in large arrays of structures [7]. Of course, it is clear that the prognosis accuracy provided by this approach is confined, on principle, by the generic features of the Hueckel model, i.e. by the topological character of the treatment and the neglect of interelectron interaction. However, it is just molecular topology that primarily contributes to the properties of a conjugated compound, while the effects of interelectronic interaction can mostly be treated as more or less considerable corrections. Such a principle justifies the prognosis of molecular behaviour based on a rather simple method like LCA. It is also of significance if the estimates given by LCA can be refined using empirical adjustment of the starting model parameters. The objective of the present study is to ascertain the real potentialities of LCA in predicting basicity and light absorption region of PMDs; we are also focused on finding the ways to improve the prognosis accuracy within the framework of the method invoked.

Here we employ the simplest LCA version which operates only on two first effective additive topological parameters accounting for the contribution of end groups to the properties of a PMD molecule. These parameters called electron donor ability  $\Phi_0$  and effective length  $L$  are determined only by the end group structure and unequivocally specify, provided the constant length of the polymethine chain (PC), the positions of the PMD frontier levels and, therefore, the dye reactivity (basicity and redox properties) as well as the position of its long-wavelength absorption maximum [4-9]. In the course of the development of the approach, the definitions of the topological parameters were changed for convenience of analytical treatment [8, 9] and the approximation itself has recently been generalized so that it can now take into account not only the frontier but also any levels of a PMD molecule [10]. In the present study, it was appropriate, however, to employ the quantities  $\Phi_0$  and  $L$  defined as in early works on LCA [4-6].

The task was to calculate the topological parameters  $\Phi_o$  and  $L$  for large arrays of PMDs with experimentally measured basicity and absorption characteristics and to derive, on this basis, the correlations between  $pK_a$  and  $\Phi_o$ , and between  $\lambda$  and  $L$ . To this end, we considered symmetric PMDs I containing heterocyclic end nuclei  $Z$  of 6 structural series; the experimental PMD characteristics were taken from the database created in FOMOS [11, 12] (see Appendix where the array of heterocyclic nuclei under study is presented along with the relevant theoretical and experimental data).



The PMD array involved in the basicity approximation included 66 compounds whereas the absorption maximum was approximated based on the data for 95 compounds. The quantities  $\Phi_o$  and  $L$  were found in terms of Green's functions of the end groups  $Z$  which were conveniently calculated as the elements of the inverse topological matrices of the  $Z$  structures [13]; the correlation equation were deduced by the least-squares method. In what follows, we discuss the nature and the features of the correlations obtained as well as the possibility to improve the  $pK_a$  and  $\lambda$  estimates.

## II. Approximation of PMD basicity and absorption region using LCA parameters

### 1. Topological basis and corresponding correlation peculiarities

Within LCA, the contribution of PMD end groups to the positions of the molecular frontier levels relative to the Fermi level is specified by the topological effective parameter  $\Phi_o$  called electron donor ability of a heterocyclic residue (see Fig. 1). The parameter used here varies within the range  $[0^\circ; 90^\circ]$ , i.e., it is defined as  $\Phi_o - (\text{sign } \Phi_o) \cdot 90^\circ$  in reference to the initially calculated alternating-sign electron donor ability [14]. The effect of heterocyclic end residues on the width of the PMD energy gap is characterized by another topological parameter,  $L$ , effective length of a heterocyclic residue (see Fig. 1).

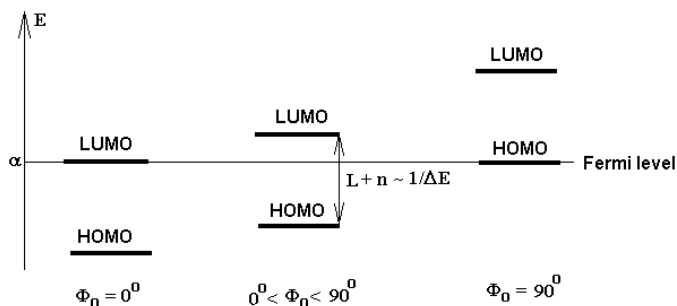


Fig. 1. The physical meaning of the parameters  $\Phi_o$  and  $L$  ( $\alpha$  is the energy of the  $2p_z(\pi)$  electron of a carbon atom,  $n$  is the number of vinylene groups in the PMD molecule).

The physical meaning of the quantities  $\Phi_0$  and  $L$  is expressed by the following relationships:

$$\Phi_0 = 90^0 \cdot \lim_{n \rightarrow \infty} [(E_{\text{HBMO}} - \alpha) / (E_{\text{HBMO}} - E_{\text{B3MO}})] \quad (1)$$

$$L = \lim_{n \rightarrow \infty} [\pi\beta / (E_{\text{B3MO}} - E_{\text{HBMO}}) - n] \quad , \quad (2)$$

where  $E_{\text{HOMO}}$  and  $E_{\text{LUMO}}$  denote the energies of the highest occupied and lowest unoccupied molecular orbitals (MOs) of a PMD molecule which are reckoned from the level  $\alpha$  (the energy of the  $2p_z(\pi)$  electron of a carbon atom) in units of  $\beta$  (the C–C bond resonance integral in a conjugated system);  $n$  is the number of vinylene groups in PC of a dye (LCA implies that  $n \rightarrow \infty$ ).

Electron donor ability and effective length are determined solely by the end group structure in the topological approximation. They are expressible in terms of the elements of the inverse topological matrices  $\mathbf{T}^{-1}$  (Green's functions) of nuclei Z [13]. If the end group Z is bound to PC through the  $t^{\text{th}}$  atom, the quantities  $\Phi_0$  and  $L$  appear as

$$\Phi_0 = -\arctg(T^{-1})_t^{-1} \quad (3)$$

$$L = [2(T^{-2})_t + 1] / [(T^{-1})_t^2 + 1] \quad (4)$$

As seen from Fig. 1 and definitions (3), (4) of the topological indices concerned, they are remarkable for an absolutely evident physical meaning; actually, they themselves are certain physicochemical characteristics expressed in terms of the simplified Hueckel model. Such an advantageous feature distinguishes them among a variety of molecular graph invariants used in establishment of structure-property relationships and enables some facilitation of the statistical work normally needed for the model to be predictive (like, e.g., thorough analysis performed in Refs. 1, 2). The values of  $\Phi_0$  and  $L$  used in further correlation procedures were found by formulae (3) and (4). The atom and bond constants involved in the topological matrices of nuclei Z are presented in Appendix.

In accordance with the known definition of effective length [4-6], the quantities  $\lambda_{\text{th}}$  for the PMDs under study were calculated as follows:

$$\lambda_{\text{th}} = (L + n) \cdot V, \quad (5)$$

where  $n = 1$  for all the dyes concerned, and  $V \approx 100$  nm is an empirically found constant, so-called vinylene shift, which characterizes an increase in the first-transition wavelength resulting from

lengthening of the PC by a vinylene group. The values of the topological parameters  $\Phi_0$  and  $L$  obtained for the Z array as well as the values of  $\lambda_{th}$  calculated for the corresponding PMDs are indicated in Appendix alongside the corresponding structures.

Since LCA does not include interelectronic interaction in PMD molecules, the model employed suggests that the frontier MOs can become infinitely close to one another, whence it follows that the first electronic transition wavelength has no upper limit as the dye colour is deepened, e.g., due to lengthening of the polymethine chromophore or binding of end groups Z with an extremely large effective length. A real bathochromic shift cannot however have an infinite magnitude: as a molecular chromophore lengthens, the energy gap tends to a constant width [4] (see Fig. 2), which becomes noticeable even within the PPP approximation [15].

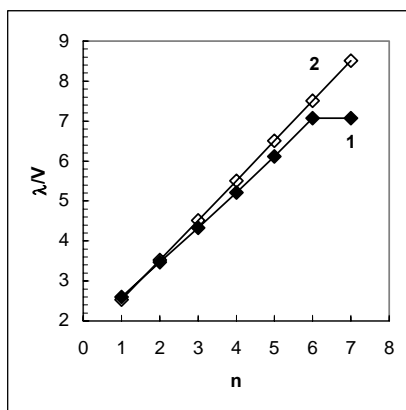


Fig. 2. The  $n$ -dependence of  $\lambda/V$  derived from experimental data (1) and calculated within LCA (2) for the PMD vinylogous series  $\text{Me}_2\text{N}^+=\text{CH}-(\text{CH}=\text{CH})_n-\text{NMe}_2$  ( $\lambda$  is the first-transition wavelength,  $V \sim 120$  nm is the vinylene shift) [4].

An impossibility to account for the nonzero limiting width of the molecular gap, an essential trend in the behaviour of molecular energy levels, leads LCA, like any other approach within the Hückel approximation, to fail in quantitative estimation of PMD properties. At the same time, this shortcoming of the method manifests itself to a different degree in the calculation of various PMD  $\pi$ -characteristics. The most plausible are those dictated by electron donor ability, i.e., the position of molecular frontier levels with respect to the Fermi level (see Fig. 1), such as the dye basicity which is symbate to the value of pKa. Indeed, the relative disposition of the molecular energy gap and the Fermi level is a so-called intensive property which appears practically constant within a given vinylene series and is only slightly sensitive to interelectronic interaction effects. As a result, the correlation between the parameter  $\Phi_0$  and the experimental values of pKa should be rather well

pronounced and the corresponding  $\Phi_0$ -dependence of pKa is expected to be described quite accurately even in the linear approximation. Another situation is realized in the estimation of the first-transition wavelength  $\lambda$ , because it is given by the reciprocal width of the PMD energy gap and strongly depends on interelectronic interaction effects. One can a priori assume that the correlation between the parameter  $L$  inversely proportional to the energy gap width (see Fig. 1) and the spectroscopic experimental data will be worse than in the case of basicity prediction. In addition, it is expedient to approximate the dependence  $\lambda$  of  $L$  by a homographic rather than a linear function, so as to reproduce a tendency to saturation observed for the first-transition wavelength in the long-wavelength absorption region. It is noteworthy that  $\Phi_0$  and  $L$  are the free term and the coefficient of the linear term, respectively, in the expansion of the end-group generating function in terms of energy, while the other effective additive parameters (not involved in this study) represent the coefficients of the quadratic and higher terms [6]. The above-mentioned regularities of the correlations based on  $\Phi_0$  and  $L$  are also valid for the correlations concerning “higher” parameters: the coefficients of even powers of energy, like electron donor ability  $\Phi_0$ , describe the contributions of the nucleus  $Z$  to intensive (independent of the PC length) molecular properties and provide reasonably good accuracy of the relevant estimates; in contrast, the coefficients of odd powers of energy, such as effective length  $L$ , characterize extensive (increasing with the PC length) molecular constants which are much more difficult to approximate within LCA.

Our present attempt to approximate PMD basicity and absorption region by the topological parameters  $\Phi_0$  and  $L$  corroborate, in full measure, the a priori assumptions on their applicability; at the same time, it affords the potentialities for notable improvement of the estimation accuracy even in the case of extensive molecular characteristics.

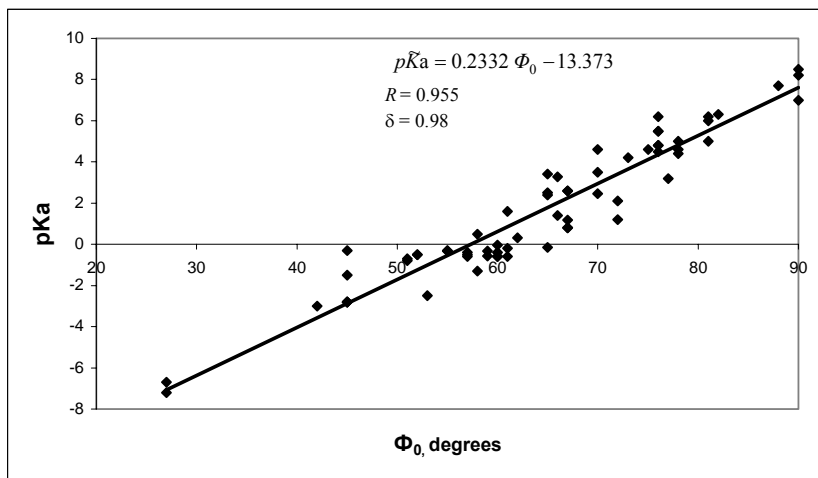
## **2. Estimation of dye basicity by the value of pKa using electron donor ability $\Phi_0$ of nucleus $Z$**

The relationship between PMD basicity and the value of  $\Phi_0$  for the nucleus  $Z$  ( $\Phi_0 = -90 \pm 90^0$ ) was found as early as in 1989 [16] for 16 nitrogen-containing heterocyclic residues widely used in the synthesis of spectral sensitizers. Finding the linear dependence pKa of  $\Phi_0$ ,  $p\bar{K}a = 0.24 \Phi_0 - 13.73$ , with the correlation coefficient  $R$  equal to 0.95, enabled calculation of pKa values for a variety of PMDs and also stimulated research on the relationship between electron donor ability and reactivity of heterocycles [17].

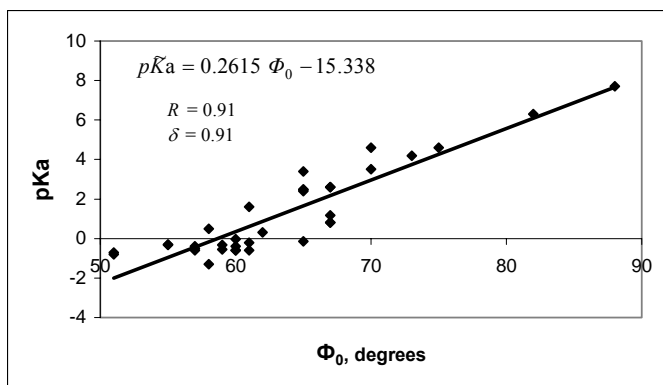
Though the present study involves an increased array of 66 nuclei  $Z$ , with the basicity range extended from low-basicity (the pyrrole series) to high-basicity (the imidazole series) PMDs, the corresponding correlation equation appears as  $p\bar{K}a = 0.23 \Phi_0 - 13.37$  ( $R = 0.96$ ), i.e., it is much the same as the previously found one. This is indicative of a rather stable dependence (see also Fig. 3a).

A small mean-square deviation of the experimental data from the approximation straight line,  $\delta = 0.96$ , points to high reliability of the estimates based on the correlation found.

On dividing the Z array into the groups of structurally related nuclei, the parameters of the correlation dependence within separate groups remain practically unchanged, though the correlation coefficient somewhat diminishes as a result of the decreased PMD sample and hence restricted  $\Phi_0$  and pKa intervals (as an example, cf. the results for the total sample in Fig. 3a and for the array concerning the thiazole series in Fig 3b).



a



b

Fig. 3. Linear approximation of measured pKa values using electron donor ability  $\Phi_0$  for (a) 66 PMDs of the full (structurally miscellaneous) array and (b) 35 PMDs of the thiazole subarray (tilded characters denote the approximated quantities, as distinct from experimental data).

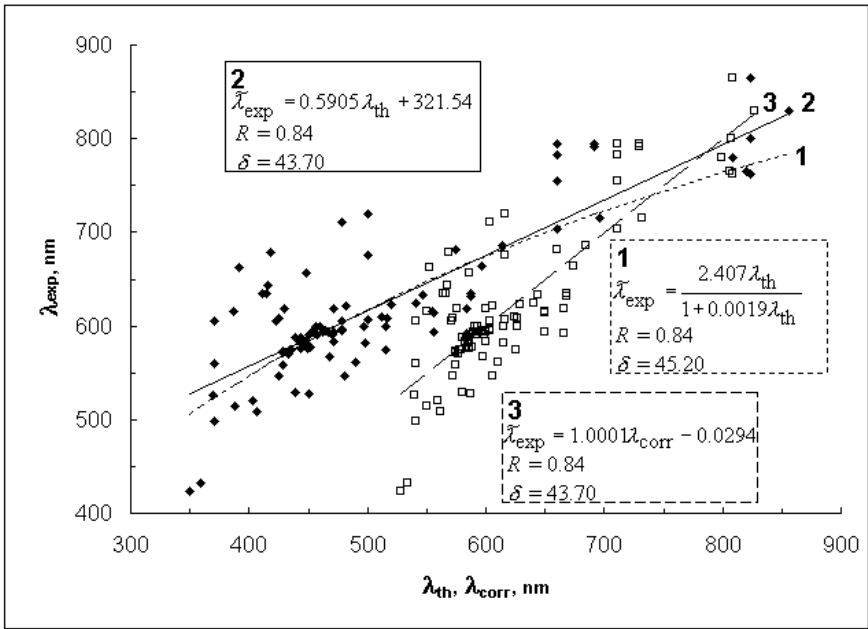
With the linear relationship between  $\Phi_0$  and  $pK_a$ , one can calculate, starting from the values of  $\Phi_0$ , the values of  $pK_a$  for new or even hypothetical (hitherto unsynthesized) PMDs as well as for the systems which do not allow direct  $pK_a$  measurements. The reasonably good correlation between  $\Phi_0$  and  $pK_a$  suggests that the PMD basicity estimates made in the framework of LCA are practically promising, e.g., applicable in the development of PMD databases [11, 12].

The correlation found is also helpful in solution of the inverse problem: it may well be necessary to estimate the contributions of various substituents to the values of  $\Phi_0$  for heterocyclic residues, given experimentally measured values of  $pK_a$  for the corresponding substituted PMDs. (It is known that the parametrization of heteroatoms within Hueckel method and hence LCA is underelaborated and in many cases indeterminate, which makes difficulties for the calculation of topological parameters for substituted nuclei. On the other hand, substituent effects are sometimes drastic and cannot therefore be neglected. To exemplify, the groups  $-\text{SO}_2\text{CF}_3$  introduced into 2-benzothiazolyl and 2-benzimidazolyl residues decrease their  $\Phi_0$  values by 8 and  $10^0$ , respectively [11, 16]).

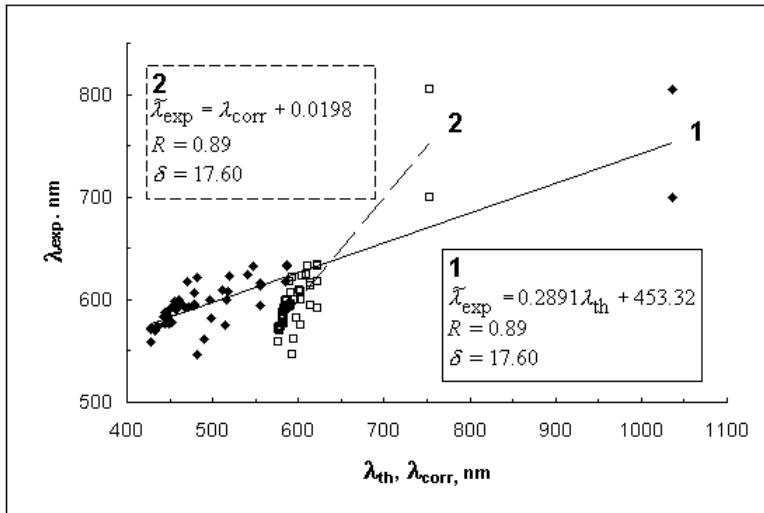
### ***3. Estimation of the wavelength of dye first electronic transition using effective length $L$ of nucleus $Z$***

The  $\lambda_{\text{th}}$ -dependence of  $\lambda_{\text{exp}}$  shows the behaviour (see Fig. 4) which confirms the expected overestimation of the calculated effective length in the long-wavelength spectral region (see also Fig. 2). Indeed, this dependence noticeably deviates from linearity at wavelengths beyond 600 nm and has a tendency to form a “plateau” at the wavelength  $\lambda_\infty$  which corresponds to a nonzero, though very narrow, gap characteristic of sufficiently long-wavelength PMDs [4,14]. To phenomenologically describe this feature of the plot, we approximate the dependence  $\lambda_{\text{exp}}$  of  $\lambda_{\text{th}}$  by a homographic function  $\tilde{\lambda}_{\text{exp}} = a\lambda_{\text{th}}/(1 + b\lambda_{\text{th}})$ , where the limiting wavelength is specified by the relation  $a/b = \lambda_\infty$  and the vinylene shift is appropriately renormalized by the factor  $a$ . For an array of 95 PMDs having various structures, we find (by the least-squares method) the approximation curve (the dotted line 1 in Fig. 4) with  $a = 2.407$ ,  $b = 0.0019$ , the correlation coefficient  $R = 0.84$ , and the mean-square deviation from the approximation  $\delta = 45.20$  nm. As easily found from this formula, the first-transition wavelength will not exceed  $\sim 1270$  nm at an arbitrarily long PC of a PMD.





a



b

Fig. 4. a: Homographic (dotted line 1) and linear approximations of the measured first-transition wavelength using effective length  $L$  (solid line 2,  $\blacklozenge$ ) and corrected effective length  $L_{corr}$  (dashed line 3,  $\square$ ) for 95 PMDs of the full (structurally miscellaneous) array; b: Linear approximation of the measured first-transition wavelength using effective length  $L$  (solid line 1,  $\blacklozenge$ ) and corrected effective length  $L_{corr}$  (dashed line 2,  $\square$ ) for 51 PMDs of the thiazole subarray (tilted characters have the same meaning as in Fig. 3).

It is informative to compare the efficiency of the homographic and the linear approximations, the latter having the form

$$\tilde{\lambda}_{\text{exp}} = a\lambda_{\text{th}} + b. \quad (6)$$

The linear correlation equation (6) with  $a = 0.5905$  and  $b = 321.54$  ( $R = 0.84$ ,  $\delta = 43.70$  nm) is represented by a dashed line in Fig. 4a. As seen, the mean-square deviation  $\delta$  of experimental points from the linear approximation (6) is even less than the corresponding value in the case of the homographic approximation, which adds to the attractiveness of the simple linear approach. Moreover, the parameters of the linear correlation equation can be used to correct the values of  $L$  so that the corresponding estimates of absorption wavelengths (calculated by Eq. (5)) more accurately reproduce experimental data. Let us consider the formula relating the mean-square variance  $\tilde{\delta}$  between experimental and calculated data to the mean-square deviation  $\delta$  of approximated from experimental values:

$$\begin{aligned} \tilde{\delta}^2 &\equiv \frac{1}{M} \sum_{i=1}^M (\lambda_{\text{exp},i} - \lambda_{\text{th},i})^2 = \delta^2 + (a-1)^2 \delta_x^2 + [(a-1)\langle x \rangle + b]^2, \\ \langle x \rangle &\equiv \frac{1}{M} \sum_{i=1}^M \lambda_{\text{th},i}, \quad \delta_x^2 \equiv \frac{1}{M} \sum_{i=1}^M (\lambda_{\text{th},i} - \langle x \rangle)^2, \end{aligned} \quad (7)$$

where  $M$  designates the number of sample points. For the selected array of 95 PMDs, the value of  $\tilde{\delta}$  amounts to 131.56 nm and so a large error makes questionable the very possibility to apply the LCA method in practically oriented estimations of absorption wavelengths. On the other hand, this value of variance is hugely contributed by the so-called systematic error of the calculation which arises from the neglect of interelectronic interaction within LCA. Graphically, the systematic error manifests itself by the deviation of the approximation straight line calculated by Eq. (6) from the bisector of the first quadrantal angle, i.e.  $a \neq 1$  and  $b \neq 0$  in Eq. (7); as a result,  $\tilde{\delta}$  can far exceed  $\delta$ . Below we suggest a straightforward method to remove the systematic error from the calculated values of effective length using the available experimental characteristics of the sample concerned. The corrected values of effective length,  $L_{\text{corr}}$ , are obtained by equating the correlation equation (6), with regard to expression (5), and the relation analogous to (5) in which  $L$  is replaced with the desired quantity  $L_{\text{corr}}$ :

$$\tilde{\lambda}_{\text{exp}} = \lambda_{\text{corr}}, \quad \lambda_{\text{corr}} \equiv (L_{\text{corr}} + n) \cdot V \quad (8)$$

The solution of the above equation appears as

$$L_{\text{corr}} = aL + (a-1)n + b/V. \quad (9)$$

The values of  $L_{\text{corr}}$  found by formula (9) (see Appendix) are determined not only by the starting theoretical parameter  $L$  but also by the parameters of the correlation equation (6); the discussed systematic error is thus avoided, as a result of the proper use of the relevant experimental data within the sample.

Fig. 4a presents experimental absorption wavelengths versus corrected estimates  $L_{\text{corr}}$  calculated by Eq. (8) – see the dashed line 3. Comparing lines 2 and 3 makes it evident that the “correction” of the initial plot reduces to rotating it through a certain angle (elimination of a so-called methodic error) and shifting along the ordinate axis (elimination of a constant error), so that the resulting plot of the function  $\lambda_{\text{exp}}(\lambda_{\text{corr}})$  practically coincides with the bisector of the first quadrantal angle and the correlation equation becomes  $\tilde{\lambda}_{\text{exp}} = 1.0001 \lambda_{\text{corr}} - 0.0294$  (a negligible deviation of the line from the bisector is due to the accumulation of rounding errors in the summation of numerical values within large arrays). A transformation of this kind exerts no effect on the scatter of points, i.e., the coefficient of correlation between calculated and experimental values as well as the mean-square deviation of experimental data from the approximation right line remain unchanged. However, the corrected calculated values reproduce experimental data much more accurately. Indeed, expression (7) now involves the parameters  $a = 1$  and  $b = 0$ , and hence  $\tilde{\delta} = \delta = 43.70$  nm, a one-third of the initial  $\tilde{\delta}$  value, since the systematic error is eliminated from the estimate of PMD absorption maximum positions (due to the implicitly included trend toward saturation observed for sufficiently long absorption wavelengths). It is clear that the corrected values of effective length are convenient to manipulate in the prediction of PMD first-transition wavelengths. Importantly, other molecular properties which need knowledge of effective length to be estimated within LCA (as, for instance, linear and nonlinear polarizabilities [18]), will also be characterized more adequately using the quantity  $L_{\text{corr}}$ .

Due to the fact that the approximation suggested covers a great number of variously structured heterocyclic nuclei, the effective length of any end residue (not necessarily included by the sample under study) can be corrected based on the same numerical correlation parameters as found above. On the other hand, a switch from a structurally miscellaneous array of nuclei to separate series which include heterocyclic systems of related structure normally offers the advantage of better correlations between calculated and experimental data and accordingly reduced mean-square deviations of approximated from experimental absorption wavelengths. To exemplify,

we again turn to the PMD series derived from thiazole: as evidenced by Fig. 4b, this subsample is characterized by a better correlation coefficient ( $R = 0.89$ ) and a smaller mean-square deviation of approximated from experimental values ( $\delta = 17.60$  nm) than for the total sample  $Z$  (cf. Fig. 4a). The correction of the calculated absorption wavelengths by the above-described technique results in a more than 7-fold reduction of the mean-square variance  $\delta^2$  between experimental and calculated data (from 126.01 to 17.60 nm). It should be noted that a heterocyclic nucleus  $Z$  has different values of  $L_{\text{corr}}$  (and  $\lambda_{\text{corr}}$ ) within a large miscellaneous array and a small structurally homogeneous subarray (compare, e.g., the corresponding values for the nuclei of the thiazole series in Appendix), since the parameter in question is not an invariant of a given structure and changes in going from one sample to another. However, its values for the same structure included in different samples differ negligibly as compared to the difference between the initial  $L$  value and any of the corrected  $L_{\text{corr}}$  values.

### III. Conclusions

This paper presents the theoretical basis and possibilities of predicting PMD molecular characteristics starting from the topological parameters of dye end nuclei in the framework of the long-chain approximation. For a large sample of PMDs (about 100 dyes with the end nuclei  $Z$  derived from 6 heterocyclic structural types, *viz.*, pyrrole, thiazole, pyridine, imidazole, pyrylium and its thia/selena analogues, and polyazaheterocycles), we have calculated the key effective topological parameters of LCA (electron donor ability  $\Phi_o$  and effective length  $L$ ) and correlated them with experimental chemical and spectral characteristics (pKa value and the first electronic transition wavelength  $\lambda_{\text{exp}}$ ).

As inferred from the analysis of a sufficiently large array of experimental data, molecular properties of intensive character, e.g., PMD basicity can be estimated much more accurately than extensive characteristics exemplified by the first transition wavelength. Accordingly, the linear approximation of basicity by the parameter  $\Phi_o$  affords a significantly larger  $\Phi_o$ -pKa correlation coefficient and a smaller mean-square deviation of experimental data from the approximation straight line than the corresponding values in the estimation of the absorption wavelength by the parameter  $L$ .

A stable correlation between the  $\Phi_o$  and pKa series has been established ( $R = 0.96$ ). The linear dependence pKa( $\Phi_o$ ) has been found for the whole basicity interval specified by 6 structural series of heterocyclic end nuclei. The mean-square deviation of the calculated from experimental data amounts to 0.98 pKa units, i.e., about 6% of the basicity variation interval in the PMD sample under study. The relationship established between the electron donor ability of nuclei  $Z$  and PMD basicity enables one to predict, to a sufficiently high accuracy, the pKa values for new dyes starting

from the  $\Phi_0$  values for their end groups as well as to solve the inverse problem (with the object of development of substituent parametrization within the LCA model).

A moderate coefficient of the correlation between  $\lambda_{th}$  calculated by  $L$  and  $\lambda_{exp}$  (0.84-0.89) suggests that the effective length of end nuclei is an appropriate quantity rather for finding trends in dye absorption maxima than for quantitative prognostication. The mean-square deviation of experimental data from the approximation straight line,  $\delta = 43.70$  nm, amounts to about 10% of the absorption wavelength variation interval for the PMDs concerned. To ascertain the possibility of more efficient first-transition wavelength estimation by the values of  $L$ , we have considered two ways of approximating the experimentally measured wavelengths, namely, with the homographic and the linear function of  $L$ . Unexpectedly, the much more straightforward linear approximation has also proved to be more accurate than that based on the homographic function; the former technique is therefore regarded as the optimal one. Moreover, a procedure has been elaborated which involves the coefficients of the linear correlation equation to correct the topological parameter  $L$  so that the neglect of interelectronic interaction within LCA is partially compensated and hence the systematic error in the prediction of PMD absorption maxima is eliminated. With the corrected values of effective length, the variance between the calculated and experimental data is reduced by more than half an order. This correction may also be helpful in the treatment of other molecular properties of PMDs related to their absorption region (for instance, linear and nonlinear polarizabilities). Thus, effective length can serve as a certain guiding line in the selection of the nucleus  $Z$  structure which should provide PMD absorption in a desired spectral region.

To conclude, the topological parameters of the LCA approach appear to be an efficient tool in primary estimation of molecular properties and selection of promising structures in large arrays of PMD end nuclei, i.e., in the stage which precedes thorough investigation of them by more precise methods.

### **Acknowledgements**

The author is thankful to the late Professor E. B. Lifshits for initiating this study and giving an access to the FOMOS database of PMD experimental characteristics. Fruitful discussions of the work with Professor V. M. Rozenbaum and Doctor M. N. Ushomirskii are also acknowledged.

## Appendix. Heterocyclic nuclei studied and their characteristics involved in correlation analysis

PMD end nuclei  $Z$  of various structure including the series of pyrrole (I), thiazole (II), pyridine (III), imidazole (IV), pyrylium and thia/selena pyrylium (V), and polyazaheterocycles (VI), with their experimentally measured characteristics (the pKa value and the first-transition wavelength  $\lambda_{\text{exp}}$ ) and calculated parameters (electron donor ability  $\Phi_o$ , effective length  $L$ , corrected effective length  $L_{\text{corr}}$ , and the respective estimates of the first-transition wavelength  $\lambda_{\text{th}}$  and  $\lambda_{\text{corr}}$ ) indicated in a column to the right of the corresponding structure as shown on the pattern below. Unavailable values are designated by dashes. For  $L_{\text{corr}}$  and  $\lambda_{\text{corr}}$  in the thiazole series, the first number characterizes the corresponding nuclei in the structurally miscellaneous array of 95 PMDs and the second one (parenthesized) refers to the same parameter determined within the thiazole subarray of 51 PMDs. The Coulomb,  $\alpha_X$ , and the resonance,  $\beta_{XX}$ , integrals for atoms and bonds are defined respectively as  $\alpha_C + h_X\beta_{CC}$  and  $\eta_{XX}\beta_{CC}$  with the commonly used  $h_X$  and  $\eta_{XX}$  values:  $h_C = 0$ ,  $h_N = 1$ ,  $h_O = 2$ ,  $h_S = 0.7$ ,  $h_{Se} = 0.5$ ;  $\eta_{CC} = \eta_{CN} = 1$ ,  $\eta_{CO} = 0.8$ ,  $\eta_{CS} = 0.4$ ,  $\eta_{CSe} = 0.2$ .

$\Phi_o$ , degrees

pKa



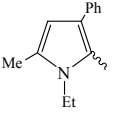
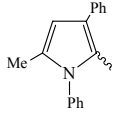
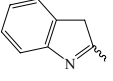
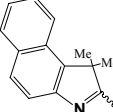

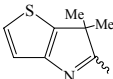
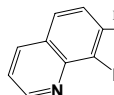
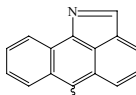
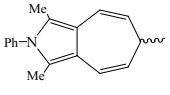

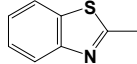
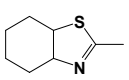
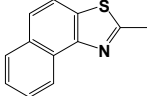
$L$

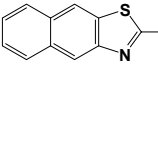
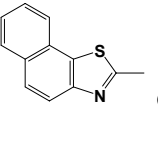
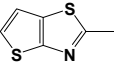
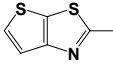
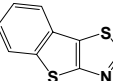
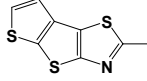
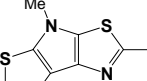
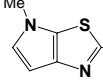
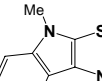
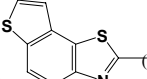
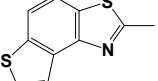
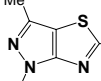

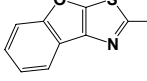
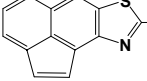
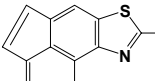
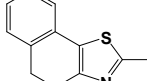
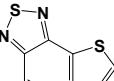
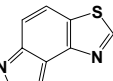
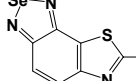
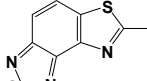
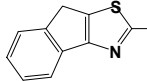
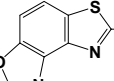
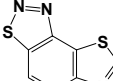
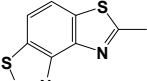
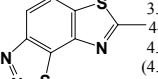
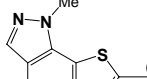
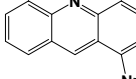
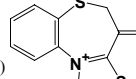
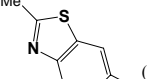
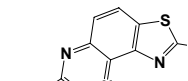
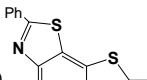
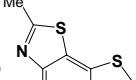
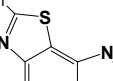
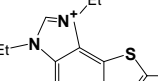
$\lambda_{\text{th}}$ , nm

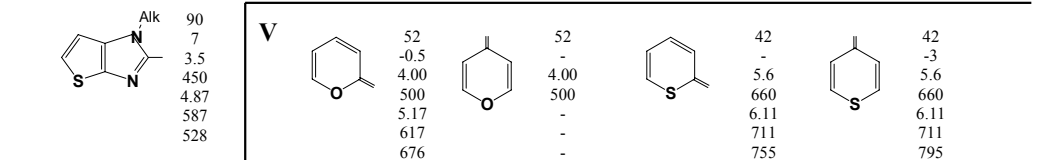
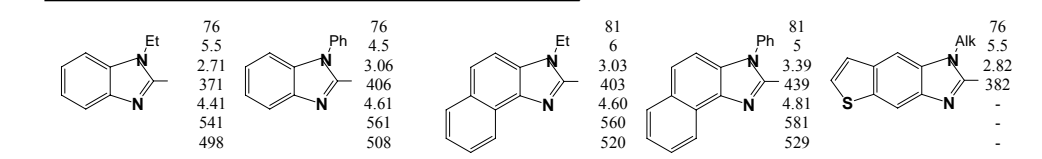
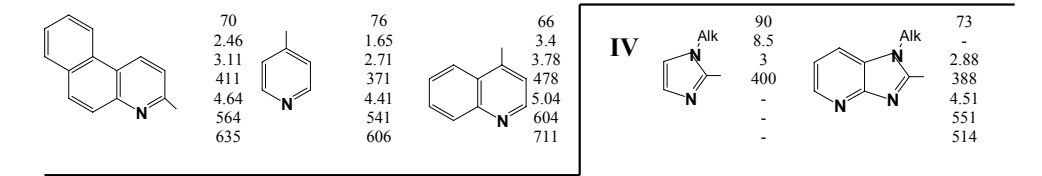
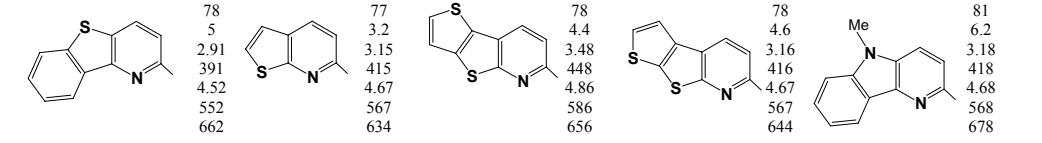
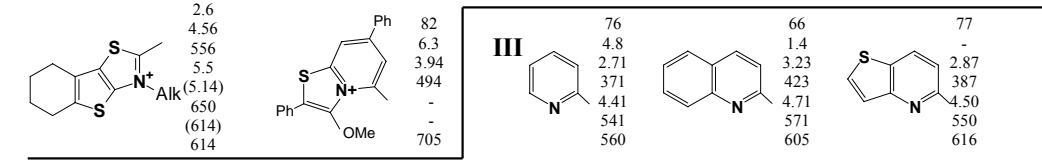
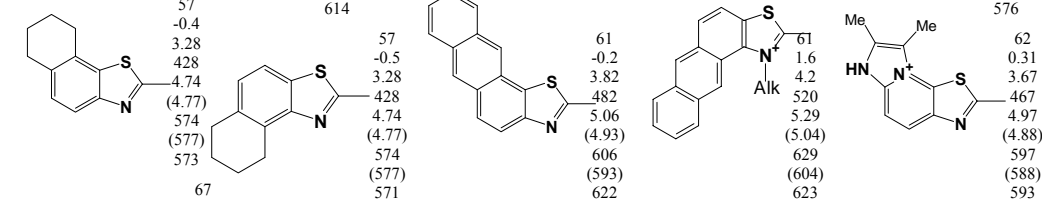
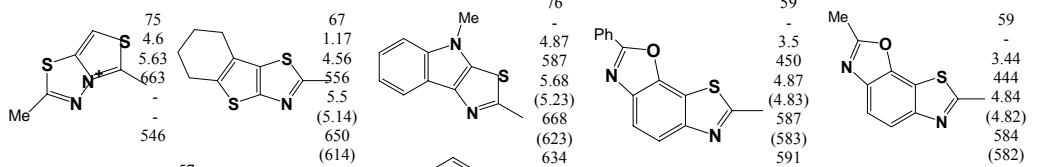
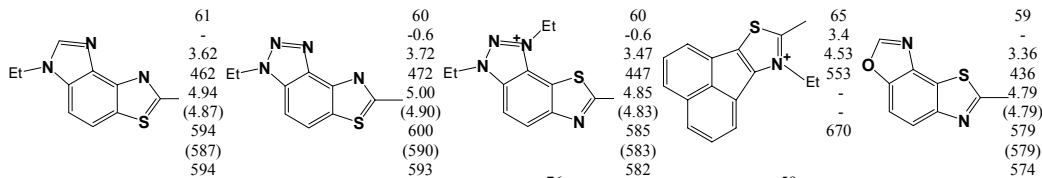
$L_{\text{corr}}$  in the whole array ( $L_{\text{corr}}$  in the subarray)

$\lambda_{\text{corr}}$  in the whole array ( $\lambda_{\text{corr}}$  in the subarray), nm

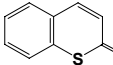
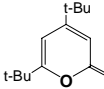
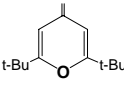
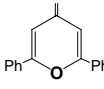
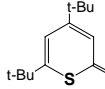
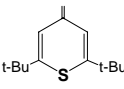
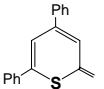
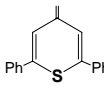
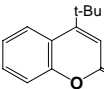
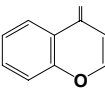
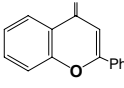
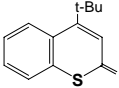
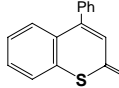
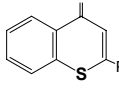
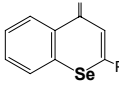
$\lambda_{\text{exp}}$ , nm

<b>I</b>	45 -0.3 2.5 350 4.28 528 424		18 - 3.7 470 4.99 599 590		18 - 3.25 425 4.73 573 608		45 -2.8 3.25 425 4.73 573 546		45 -1.5 3.39 439 4.81 581 588
	45 -2.8 6.5 750 - 691		50 - 3.68 468 4.98 598 567		45 -2.8 3.72 472 5.00 600 584		27 -6.7 7.2 820 7.06 806 765		53 -4.8 580 - - 810
<b>II</b>	51 -0.7 2.46 346 - - 444		65 2.4 3.81 481 5.06 (4.92) 606 (592) 546		57 -0.6 3.28 428 4.74 (4.77) 574 (577) 559		51 -0.8 2.46 346 - -		60 -0.04 3.78 478 5.04 (4.92) 604 (592) 595





	55 -0.3 433 4.77 (4.79) 577 (579) 570		60 -0.4 3.53 453 4.89 (4.84) 589 (584) 593		67 0.8 4.56 556 5.5 (5.14) 650 (614) 594		67 0.8 3.98 498 5.16 (4.97) 616 (597) 582		65 -0.14 4.41 541 5.41 (5.10) 641 (610) 625
	70 4.6 4.84 584 5.66 (5.22) 666 (622) 618		73 4.2 4.55 555 5.49 (5.14) 649 (614) 616		70 3.5 4.04 504 -		76 - 4.87 587 5.68 (5.23) 668(623) 632		59 -0.56 3.42 442 4.83 (4.81) 583 (581) 584
	59 -0.32 3.65 465 4.96 (4.88) 596 (588) 595		64 - 3.46 446 4.84 (4.82) 584 (582) 585		88 7.7 9.36 1036 - - -		70 - 4.17 517 5.27 (5.03) 627 (603) 608		46 - 4.15 515 5.26 (5.02) 626 (602) 575
	51 - 4.16 516 5.26 (5.02) 626 (602) 600		65 2.5 4.47 547 5.45 (5.11) 645 (611) 633		61 - 3.6 460 4.93 (4.86) 593 (586) 600		61 - 3.78 478 5.04 (4.92) 604 (592) 596		62 - 3.71 471 5.00 (4.89) 600 (589) 618
	62 - 3.78 478 5.04 (4.92) 604 (592) 606		65 3.9 490 5.11 (4.95) 611 (595) 562		59 - 3.52 452 4.88 (4.84) 588 (584) 578		58 -1.3 3.33 433 4.77 (4.79) 577 (579) 570		52 - 4.84 584 5.66 (5.22) 666 (622) 592
	58 - 3.49 449 4.87 (4.83) 587 (583) 577		60 - 3.57 457 4.91 (4.85) 591 (585) 590		55 -0.32 4.12 512 5.24 (5.01) 624 (601) 610		67 2.6 5.61 661 - - 582		58 - 3.34 434 4.78 (4.79) 578 (579) 574
	58 0.5 3.97 497 5.15 (3.97) 615 (497) 600		59 - 3.56 456 4.91 (4.85) 591 (585) 599		59 - 3.48 448 4.86 (4.83) 586 (583) 585		59 - 3.44 444 4.84 (4.82) 584 (582) 588		61 -0.6 3.49 449 4.87 (4.83) 587 (583) 576





	27 -7.2 5.91 6.91 6.30 730 795		52 - 4.00 500 5.17 617 720		52 - 4.00 500 5.17 617 607		52 - 5.14 614 5.84 684 686		42 - 5.6 660 6.11 711 782
	42 - 5.6 660 6.11 711 703		42 - 7.24 824 7.08 808 865		42 - 7.24 824 7.08 808 762		36 - 4.74 574 5.6 660 682		36 - 4.97 597 5.74 674 664
	36 - 5.96 696 6.33 733 715		27 - 5.91 691 6.30 730 792		27 - 7.09 809 6.99 799 780		27 - 7.23 823 7.08 808 800		10 - 7.56 856 7.27 827 830

**VI**

	72 1.2 3.3 430 4.75 575 618		72 2.1 2.7 370 4.40 540 526		90 8.2 2.5 350 - - -		76 6.2 2.59 359 4.34 534 432
---	---	---	---	---	--	---	--

## References

- [1] I. I. Baskin, V. A. Palyulin and N. S. Zefirov, *J. Chem. Inf. Comput. Sci.* 37 (1997) 715.
- [2] M. I. Skvortsova, K. S. Fedyaev, V. A. Palyulin and N. S. Zefirov, *Internet Electron. J. Mol. Des.* 2 (2003) 70 [<http://www.biochempress.com>].
- [3] E. B. Lifshits, R. D. Raikhina, L. G. Kurkina and M. N. Ushomirskii, *Zh. Nauch. Prikl. Fotogr. [Sci. Appl. Photo.]* 41 (1996) 43.
- [4] G. G. Dyadyusha and A. D. Kachkovskii, *Ukr. Khim. Zh.*, 41 (1975) 1176 [*Sov. Prog. Chem.* 41, No. 11 (1975) 52].
- [5] G. G. Dyadyusha and A. D. Kachkovskii, *Ukr. Khim. Zh.* 44 (1978) 948 [*Sov. Prog. Chem.* 44, No. 9 (1978) 50].
- [6] G. G. Dyadyusha and M. N. Ushomirskii, *Teoret. Eksperim. Khim.* 21 (1985) 268 [*Theor. Exp. Chem.* 21 (1985) 257].
- [7] M. Yu. Kornilov, G. G. Dyadyusha, V. I. Zamkovoii, M. L. Dekhtyar and A. D. Kachkovskii, *Khim. Geterotsikl. Soed. [Chem. Heterocycl. Comp.]* No. 2 (1984) 217.
- [8] G. G. Dyadyusha, V. M. Rozenbaum and M. L. Dekhtyar, *Zh. Eksp. Theor. Fiz.* 100 (1991) 1051 [*Sov. Phys. JETP* 73 (1991) 581].
- [9] M. L. Dekhtyar, *Dyes and Pigments* 28 (1995) 261.
- [10] M. Dekhtyar, V. Rozenbaum and S.-H. Lin, *MATCH Commun. Math. Comput. Chem.* 47 (2003) 71.
- [11] E. B. Lifshits, M. N. Ushomirskii, R. D. Raikhina and L. G. Kurkina, *Proc. VI Symp. on Physics and Chemistry of Polymethine Dyes, 1996, Moscow*, p. 117.
- [12] E. B. Lifshits, *Proc. VII Symp. on Physics and Chemistry of Polymethine Dyes, 1999, Moscow*, p. 9.
- [13] G. G. Dyadyusha, M. L. Dekhtyar and M. Yu. Kornilov, *MATCH Commun. Math. Comput. Chem.* 23 (1988) 21.
- [14] G. G. Dyadyusha and M. N. Ushomirskii, *Teoret. Eksperim. Khim. [Theor. Exp. Chem.]* 22 (1986) 137.
- [15] G. G. Dyadyusha and I. V. Repyakh, *Ukr. Khim. Zh. [Sov. Prog. Chem.]*, 57 (1991) 1155.
- [16] E. B. Lifshits, Yu. M. Karpova and M. N. Ushomirskii, *Khim. Geterotsikl. Soed. [Chem. Heterocycl. Comp.]* No. 9 (1989) 1278.
- [17] M. N. Ushomirskii, *Teoret. Eksperim. Khim.* 26 (1990) 275 [*Theor. Exp. Chem.* 26 (1990) 255].
- [18] M. L. Dekhtyar and V. M. Rozenbaum, *J. Phys. Chem. A* 101 (1997) 809.