

## BOND LENGTH IN BENZENOID HYDROCARBONS

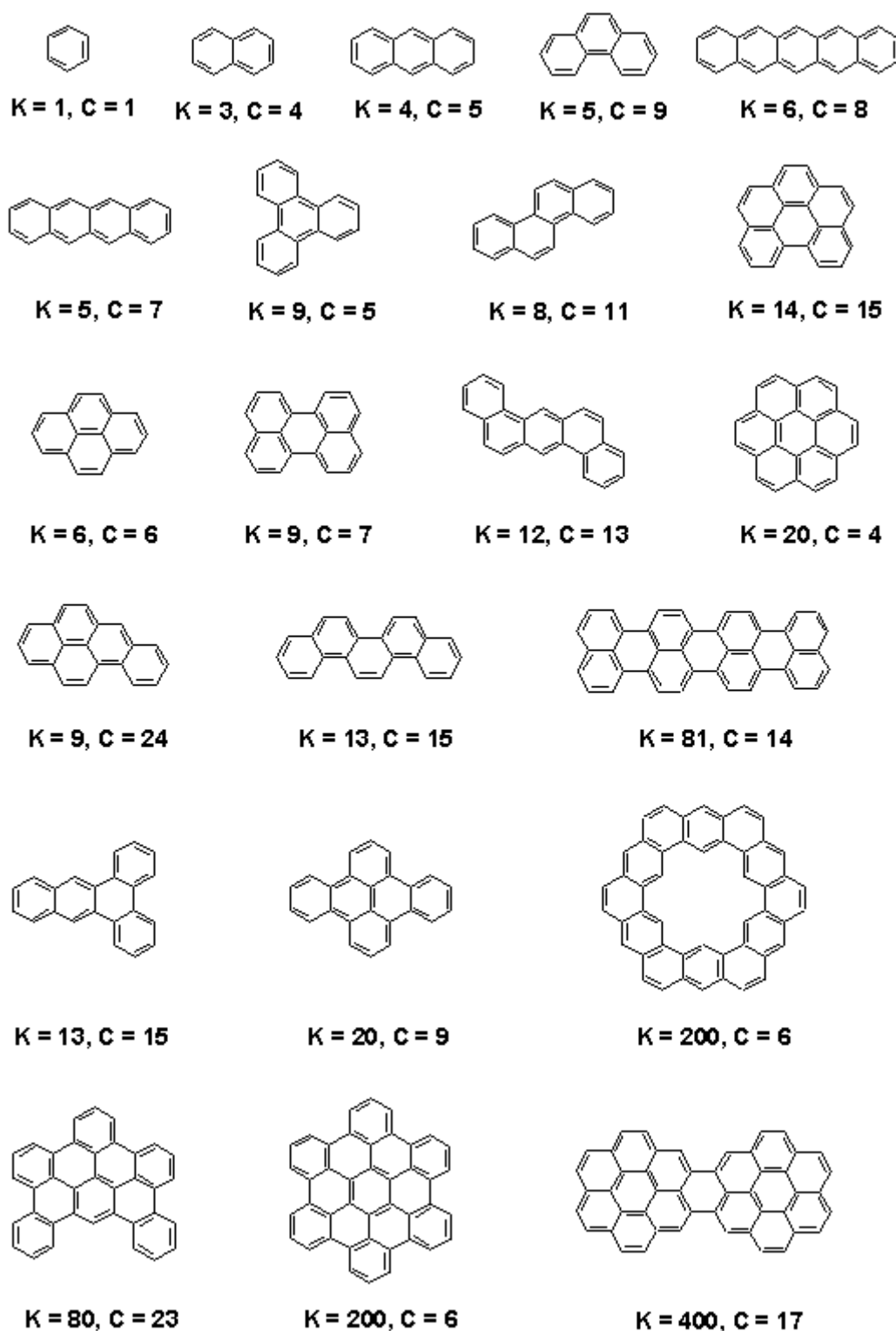
Márcia M. C. Fereira and Rudolf Kiralj, Universidade Estadual de Campinas, Campinas, SP, 13084-971, Brazil

### INTRODUCTION

Empirical calculation of carbon-carbon bond length in organic compounds, especially in aromatic hydrocarbons, is a traditional subject of structural, theoretical and organic chemistry. For example, bond length is an important index of the aromatic character of the bond, of its corresponding ring or molecular fragment, or of the whole molecule [1-5]. Furthermore, the bond length can be used for calculation of the charge in donor-acceptor complexes and other quantities [6,7]. Recent researches confirm mutual dependence of the covalent bond length and the hydrogen bond distance in crystalline state of organic compounds [8,9]. The problem of the bond length prediction is today interesting also in area of chemistry which deals with mathematical relations between the molecular structure of a compound and its physico-chemical properties (QSPR, Quantitative Structure-Property Relationship) or biological activity (QSAR, Quantitative Structure-Activity Relationship), as is illustrated in some papers [10-14]. Theoretical approach for bond length prediction by means of statistical methods, specially chemometrics, requires a set of experimental values of bond lengths and other molecular parameters. Once these data are collected, it is possible to construct a mathematical model and apply to a case of a compound without known structure. Detailed description of a complete procedure follows in the text. One can ask the following: what is really the chemical bond in an aromatic molecule, and how its bond length can be calculated in an easy and fast way, without performing quantum-chemical and other complicated calculations?

To describe what has been done in terms of the mathematical relations which enable C-C bond calculation in planar benzenoid hydrocarbons, we must first to explain concepts like benzenoid system, data mining and chemometric regression methods.

A benzenoid system or simply benzenoid [15-17] is a delocalized  $\pi$ -electronic hydrocarbon system of benzene rings, which are mutually bound directly (fused) or indirectly (through a bond or group). In Figure 1 are the formulas of some benzenoid hydrocarbons with full name Planar Benzenoid Fused Polycyclic Aromatic Hydrocarbons (PB-PAHs) [18]. The long name says almost all: that they are fused, contain two or more benzene hexagons (except benzene which is included due to statistical reasons), they are aromatic and planar (all the atoms are in the same plane).



**Figure 1.** There are twenty-two planar benzenoid hydrocarbons with one to fifteen hexagons, extracted from Cambridge Structural Database (version October 2001). The crystal structures of these aromatic systems satisfy strict conditions for data mining of crystallographic data. Each hydrocarbon is presented by one Kekulé structure, the number of Kekulé structures (K) and the number of symmetrically independent types of chemical bonds (C).

Data mining or “digging” in a data base is a searching and extraction of important information for research, that requires skill and persistency. To establish a model to calculate bond length in a benzenoid

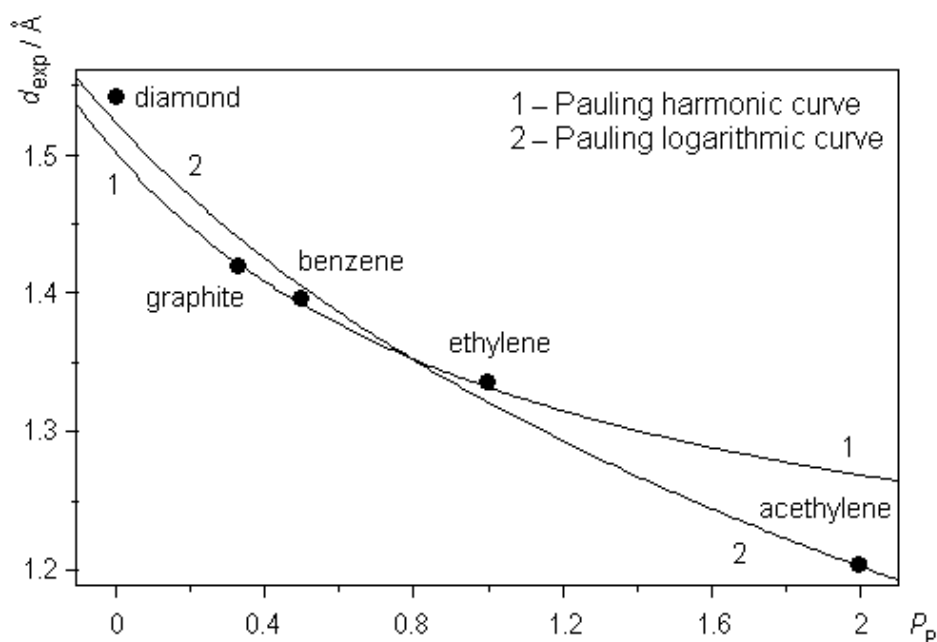
hydrocarbon, we need experimental bond lengths determined with high accuracy. The commercial Cambridge Structural Database (CSD) [19,20] contains over 257 000 crystal structures of organic and organometallic compounds (version October 2002), from which 13% have metal-ligand  $\pi$  bonds. In this base it can be found approximately hundred different benzenoid systems, what is a very small percentage of a thousand known ones [15]. When the strict criteria of quality of the determined crystallographic structures are applied for data mining for benzenoids [18,21-23], many crystal structures must be thrown away.

Chemometrics [24-39] is a relatively new chemical area for analysis of chemical data, which has roots in statistical and mathematical methods applied in psychology, sociology and technics. Although it is widely used in instrumental analytical chemistry, more and more penetrates also into theoretical chemistry, QSAR/QSPR [27,28,30-38], and all other branches of pure and applied chemistry and chemical technology. In what chemometrics can help about our concrete problem, are the regression methods. In this sense, the bond length should be considered as a dependent variable  $y$  which is a linear function  $f$  of one or more independent variables  $x_1, x_2, x_3, \dots$  etc., what can be writtein in the general form  $y = f(x_1, x_1, x_3, \dots)$ . The independent variables are called descriptors, or simply variables. About them later on in the text. Let's mention the three mostly used regression methods, which result in regression coefficients  $b_i$  in a linear expansion  $y = b_0 + b_1 x_1 + b_2 x_1 + b_3 x_3 + \dots$  Multiple Linear Regression (MLR) [27,29,31] is an inverse least squares method, an extension of usual linear regression. Principal Component Regression (PCR) [24,25,26,27,29] takes into account only that part of the original data which contains significant information, and treats it as MLR, while the rest is considered to be useless redundancy. Partial Lest Squares (PLS) [24-27,29-31] as well as PCR extracts only the relevant information, but with the difference with respect to PCR in finding the best solution by an iterative procedure on all variables, dependent and independent ones. Both methods, PCR and PLS are based on Principal Component Analysis (PCA) [24-38], which produces new descriptors as linear combinations of the original ones. In this, the new descriptors – Principal Components (PCs) are mutually orthogonal what was not the case with the original descriptors, and they are arranged in descreasing order of percentual content of the original information. The Principal Components with small percentage are thrown away as not important. More details on the mentioned methods one can find in the recommended chemometric literature.

## HISTORY OF BOND LENGTH – BOND ORDER RELATIONSHIP IN BENZENOID HYDROCARBONS

As there were crystal structures of many compounds determined by X-ray diffraction in 1930s, it was possible to notice relations between various molecular characteristics for some class of compounds. One such relation is bond length – bond order, as it is illustrated in **Figure 2** for the simplest C-C bonds [39]. In general, the relation is not linear, what can be proved today by quantum-mechanical *ab initio* calculations [3,40]. Nobel Prize Winner Linus Pauling (1901-1994) used the bond number (the bond multiplicity) as the bond order [16,41,42]. He has defined  $\pi$  bond order also (the Pauling  $\pi$  bond order,  $P_p$ ), which is equal to the

bond number reduced by one. [16,17,39,41].  $P_p$  is strictly defined as the ratio of the total number of Kekulé structures  $K$  to the number of Kekulé structures  $K_d$  in which the bond under study appears as formal double. With the increase of the number of benzenic hexagons in a molecule of a benzenoid, the number of symmetrically independent bonds does not increase necessarily (the reason for this is high molecular symmetry), but certainly  $K$  increases as well as the numbers  $K_d$  for many bonds. Pauling stated [42] that all Kekulé structures of a molecule can be drawn on a piece of paper for a few minutes up to an hour or two for the most complicated systems as the largest ones in **Figure 1**. But who has no skill in this, and does not recognize the symmetrically repeated Kekulé structures (which are in fact one and the same structure), can make a mistake easily. There are also other methods for determination of  $P_p$ , as for example the Randić procedure [16,17,43], but they are also not more efficient than the Pauling's. It would be the safest and the best way to use some program which calculates the  $P_p$  numbers. Someone can make a question: why then to calculate  $P_p$  when a molecule can be nicely modeled by a personal computer and its geometry optimized inside a few minutes by molecular mechanics or semi-empirical methods? Both these methods do not give good agreement with the experiment because they are not parametrized for our molecules but for some general set of molecules. The best are *ab initio* methods, but they are slow – the calculation can take from a few hours to some days, and even longer. It is necessary to test these methods for the proposed set of molecules and see which method and under which conditions is the best for the calculation of C-C bond lengths. Who does not have a quantum-chemical program, or is interested in ten or more molecules, or need to calculate only a few C-C bond lengths, can easily and fastly use an analytical equation of the type bond length-bond order.

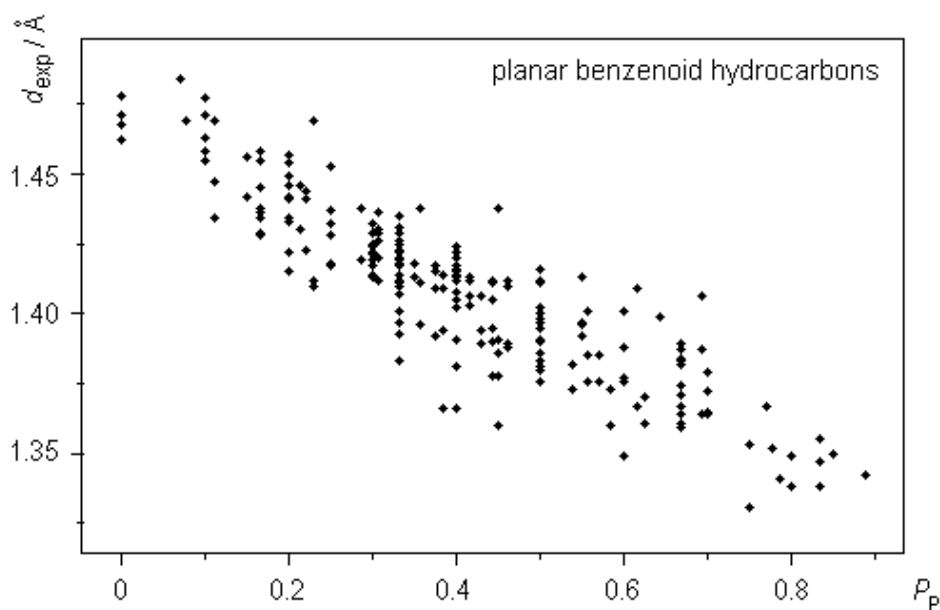


**Figure 2.** Diagram of the dependence of experimental C-C bond length on the Pauling  $\pi$  bond order for the simplest systems. It can be noticed easily that this dependence is not linear. It seems that the Pauling regression curve (logarithmic) better fits to the points than the theoretical curve of the harmonic potential, but none of them is perfect.

In the 1930s, when the principles of resonance theory were established, bond lengths were determined by diffraction techniques with accuracy at the second digit in Å units ( $\text{\AA} = 10^{-10} \text{ m}$ ). Pauling assumed that a delocalized C-C bond is a mixture of a single and double bonds, each of them behaving as a stretching harmonic oscillator. He derived theoretically [41,42] the curve of the harmonic potential for delocalized bond:  $d = d_{\text{sing}} - 1,84 (d_{\text{sing}} - d_{\text{doub}}) P_{\text{p}} / (0,84 P_{\text{p}} + 1)$  where  $d$ ,  $d_{\text{sing}}$  and  $d_{\text{doub}}$  are the lengths of delocalized, standard single and double C-C bonds, respectively. He was also the first one who used linear regression in logarithmic form [44]:  $d / \text{\AA} = a + b \log(P_{\text{p}} + 1)$ . Approximately at the same time Gordy [41] defined another type of regression for bond order prediction from the C-C bond length:  $P = a + b / d^2$ , where  $P$  is a general type of bond order, and  $d$  is a general bond length. All these equations, including the inverse Gordy equation (where the bond length can be obtained from the bond order), can be found in various recent papers [3,4,7,16-18,21-23,36,46-51]. These papers differ from the original ones of Pauling and Gordy in that the number and the accuracy of crystal structures is greater, so the regression coefficients are different. The set of hydrocarbon molecules or carbon species as grafites or fullerenes which can be found in the CSD is growing continuously.

In the 1970s appear the works of Herndon [50], Herndon and Párkányi [51] and Pauling [42] with larger sets of more accurate aromatic structures. Herndon used linear regression which had been already introduced by Pauling [39]. In 1990s our works were published in order to reparametrize this linear regression [18,21-23,46]. Data mining in the CSD database was performed, with limitation to a more homogeneous set of systems – the structures of planar benzenoid hydrocarbons with high accuracy. Bonds in non-planar benzenoids, especially in helicenes, are significantly influenced by intramolecular steric interactions, so they should be studied as a separate subject. Herndon and Párkányi had 13 benzenoid systems with a total of 100 symmetrically independent bonds (including non-benzenoids as butadien and graphite) [51], and Pauling only 9 sytems with 82 bonds [41]. Our data mining resulted at first time in 14 systems and 124 bonds in 1995 [21], then in 16 systems and 147 bonds in 1996 [22], and in 22 systems and 223 bonds in 2000 [18]. This last set is illustrated in **Figures 1** and **3**. As the accuracy of crystallographically determined bond lengths has been increased a few times with respect to that of 1930s, many effects have been discovered as for example: steric effects and the effects of intermolecular interactions on molecular structure, then crystal packing effects on molecular symmetry and geometry, substitution, isomerism and complexation effects on the geometry of molecular skeleton. The shortening of the bond lengths in crystal was noticed also, what is in fact caused by vibrations in crystal and a disadvantage of the X-ray diffraction (which locates centroids of the electron cloud instead of atomic nuclei). The relation bond length – bond order in **Figure 3** is linear or curvilinear, being hard to decide which fits better due to pronounced dispersion of the data and narrow interval for  $P_{\text{p}}$ . Our investigations confirmed that this relation is not better in statistical sense if having more molecules *i.e.* bonds, and the same is about compounds similar to benzenoids and which have aromatic C-C bonds [21-23,46]. It is worth to say that there are authors which instead of Pauling  $\pi$  bond order use Coulson, Hückel (from HMO calculations) and SCF (Self Consistent Field, from MO calculations)

bond orders [47,51]. There is high correlation between all these  $\pi$  bond orders, so none is preferred for bond length prediction.



**Figure 3.** Diagram of the dependence of experimental C-C bond length on the Pauling  $\pi$  bond order for twenty-two planar benzenoid hydrocarbons from the CSD base of crystallographic data. It is hard to judge if the points are grouped along a line or a curve because the dispersion is high, and the interval for the bond order is just from 0 – 1.

In 1960s Stoiceff developed another approach for empirical prediction of a general C-C bond length, based on structural data from spectroscopy of small molecules [52]. He obtained an equation in the form  $d / \text{\AA} = 1.299 + 0.040 n$  where  $n$  is the number of neighbouring carbon atoms which are directly bound to the carbons of the bond under consideration. In our interpretation of crystallographically determined bond lengths, this equation should be  $d / \text{\AA} = 1.315 + 0.0384 n$ . In Table 1 are the experimental [53] and calculated bonds of formal single bonds. One should have in mind the hybridization of a carbon atom which determines the number of the neighbouring atoms, as well as the fact that every bond appears as formal single in some Kekulé structures.

Herdon and Párkányi [51] and Dewar and Gleicher [54] have pointed out that the length and the order of a chemical bond are not the only phenomena mutually related when the subject is the chemical bond. In our work [18] it was proved that the dispersion of points in **Figure 3** is in accord with this fact, or by other words, it is necessary to include other quantities in the mathematical equation. Some other investigations led to the same conclusion [22,46]. Very high correlation coefficient between experimental and calculated bond lengths is not always a good measure of the predictive quality of a mathematical model. In chemometrics, there are also other statistical parameters for judging the model's validation, which are more sensitive than the correlation coefficient (see the recommended literature for this area).

**Table 1.** Experimental ( $d_{\text{exp}}$ ) and calculated ( $d_{\text{cal}}$ ) lengths of formal single C-C bonds (in units  $\text{\AA}=10^{-10}$  m) and deviation  $\Delta = d_{\text{exp}} - d_{\text{cal}}$  for different chemical neighbourhood  $n$ . The linear regression equation has form  $d / \text{\AA} = 1.315 - 0.0384 n$ . The correlation coefficient is  $R = 0.967$ , and the mean absolute  $\Delta = 0.009 \text{\AA}$ , what is satisfactorily good with respect to the small data set. The carbon hybridization types are marked by indices: sp<sup>3</sup>, sp<sup>2</sup>, sp and ar (aromatic). To obtain the number  $n$ , it is necessary to perceive how many bonds on a carbon atom are still free taking into account its hybridization (1, 2 or 3).

Bond	$n$	$d_{\text{exp}} / \text{\AA}$	$d_{\text{cal}} / \text{\AA}$	$\Delta / \text{\AA}$
C <sub>sp<sup>3</sup></sub> – C <sub>sp<sup>3</sup></sub>	6	1.530	1.545	-0.015
C <sub>sp<sup>3</sup></sub> – C <sub>sp<sup>3</sup></sub> = C	5	1.507	1.507	0
C <sub>sp<sup>3</sup></sub> – C <sub>ar</sub>	5	1.513	1.507	0.006
C <sub>sp<sup>3</sup></sub> – C <sub>sp</sub> $\equiv$ C	4	1.466	1.469	-0.003
C = C <sub>sp<sup>2</sup></sub> – C <sub>sp<sup>2</sup></sub> = C	4	1.460	1.460	-0.009
C = C <sub>sp<sup>2</sup></sub> – C <sub>ar</sub>	4	1.483	1.469	0.014
C <sub>ar</sub> – C <sub>ar</sub>	4	1.490	1.460	0.030
C = C <sub>sp<sup>2</sup></sub> – C <sub>sp</sub> $\equiv$ C	3	1.431	1.430	0.001
C <sub>ar</sub> – C <sub>sp</sub> $\equiv$ C	3	1.434	1.430	0.004
C $\equiv$ C <sub>sp</sub> – C <sub>sp</sub> $\equiv$ C	2	1.377	1.390	-0.013

## THE CHEMOMETRIC APPROACH IN SOLVING THE PROBLEM

Our chemometric approach to find an equation which would be simple and in good agreement with experimental bond lengths considered the bond length as a function of more variables [18], analogously to a QSAR or QSPR investigation [27,28,30-38]. Besides  $P_p$  other bond descriptors were defined:  $P_{\text{cr}}$  – a modified  $P_p$  containing the crystal packing effects on C-C bond lengths (for details see ref. [18]), the number  $n$ , the number  $m$  – the total number of hexagons containing the carbons of a particular bond, and the number  $l$  – the number of carbon atoms which are directly bound to those atoms counted to obtain  $n$ .

Furthermore, MLR, PCR and PLS regression models were used for the set of descriptors ( $P_p$ ,  $P_{\text{cr}}$ ,  $n$ ,  $m$ ,  $l$ ) [18]. The regression results are compared with those from simple linear regression (LR), what can be seen in Table 2 where are some statistical parameters. It should be said that the Pauling harmonic and logarithmic equations, as well as the inverse Gordy equation, produced worse results than the LR.

Besides already defined  $R$  and  $\Delta$ , the other parameters are:  $Q$  – the correlation coefficient from validation, SEV – the Standard Error of Validation, and parameter  $T$  defined by the expression  $T = \langle |d_{\text{exp}} - d_{\text{cal}}| / \sigma \rangle$  where  $\sigma$  is the experimental error of the bond length. The validation performed in our work [18] is a procedure where the bond length is obtained from the regression model which did not include this bond. Details about validation can be found in the recommended chemometric literature [24-28], as well as in our last works [18,30,31,46]. Validation of a regression model is very much important because it reveals if the models with high  $R$  and low  $\Delta$  are really good. It can be noticed in Table 2 that the PLS model II with two

Principal Components (2PC) is good enough for our practical needs. Inclusion of all 5PC (such PLS or PCR model is identical to the MLR model IV) does not give better model. On another side, simple models with as least as possible number of selected descriptors and PCs, are preferred. The models with descriptors ( $n, m, l$ ) obviously could be better explored. Therefore, at this stage of research, it is necessary to include the total information contained in these descriptors, so we recommend MLR model VI for practical purposes.

**Table 2. Significant regression models for C-C bond length calculation in planar benzenoid hydrocarbons. The best models I, II and VI are given in the form of equation which is practical even for a pocket calculator. An exceptionally good model should be characterized by high  $R$ ,  $Q > 0.90$ , small SEV,  $\Delta < 0.005 \text{ \AA}$  and  $T < 2.58$ . The average experimental error for the set from Figure 1 is  $0.005 \text{ \AA}$ .**

No.	method	descriptors	$R$	$Q$	SEV / $\text{\AA}$	$\Delta$ / $\text{\AA}$	$T$
I	LR	$P_P$	0.895	0.798	0.014	0.011	3.34
II	PLS (2PC)	$P_P, P_{cr}, n, m, l$	0.940	0.938	0.011	0.007	2.18
III	PCR (4PC)	$P_P, P_{cr}, n, m, l$	0.940	0.937	0.011	0.007	2.19
IV	MLR	$P_P, P_{cr}, n, m, l$	0.958	0.955	0.009	0.006	1.83
V	PLS (2PC)	$n, m, l$	0.820	0.814	0.018	0.014	4.60
VI	MLR	$n, m, l$	0.837	0.830	0.017	0.014	4.49
Equation I: $d / \text{\AA} = 1.468 - 0.147 P_P$							
Equation II: $d / \text{\AA} = 1.431 - 0.060 P_P - 0.063 P_{cr} + 0.006 n + 0.004 m + 0.001 l$							
Equation VI: $d / \text{\AA} = 1.464 + 0.082 n + 0.072 m - 0.094 l$							

After obtaining the regression equations I, II and VI, we can illustrate their practical side by an example. Of course, this example can be generalized to the whole set of benzenoids from **Figure 1**. Benzenoid hydrocarbon 3,4-benzopyrene is the first from left, the fourth row in **Figure 1**, and has the largest number C of bonds. There are two important factors when comparing our equations with quantum-chemical calculations: CPU time and agreement with experiment. Molecular mechanics calculations with force fields MMFF94 and SYBIL, semi-empirical PM3, AM1, MNDO, as well as *ab initio* calculations HF, DFT (SVWN and B3LYP) with the basis set 6-31G\*\* were performed by using program Titan program [55] for molecular graphics and modeling at personal computer Pentium III (330 MHz, 128 KB memory). Bond descriptors ( $n, m, l$ ) were calculated on the basis of the two-dimensional formula from **Figure 1**. Bond order  $P_P$  was determined by Pauling count of Kekulé structures, and  $P_{cr}$  according to the rules for crystal packing effects [18]. When talking about CPU time, *ab initio* varied from 100-160 minutes, semi-empirical 7-15 minutes, and the molecular mechanics methods only 2 – 4 minutes. On the other side, calculation of the bond descriptors and use of analytical equations lasted 15 – 30 minutes, and if using a pocket calculator it would be necessary 10 – 20 minutes more. When considering the quality of the models, comparison of all calculations was performed in two ways. In one case, a data matrix whose elements are bond lengths, was



constructed. Each column contained numbers obtained from the same method, including the experiment, so that the matrix dimensions are 24x12. The matrix was then treated by chemometric classificatory methods PCA and HCA (Hierarchical Cluster Analysis) [24-27,30-38]. In another case of the comparison of all results, the statistical parameters  $R$ ,  $\Delta$  and  $T$  were calculated, and the matrix of dimensions 11x3 was constructed for the identical PCA-HCA procedure. Details on this type of analysis can be found in our last works [18,46]. Both comparisons showed the following: it is clear that the *ab initio* methods are the best, then follows equation II. The third place is shared between equation I and semi-empirical methods PM3 and AM1. At the fourth position is equation VI and MNDO. The methods of molecular mechanics are the worst ones.

The equation II has perspective even with respect to *ab initio* calculations. The CPU time of a quantum-chemical calculation increases with the third power of the number of atoms. The number of C-C bonds as well as the number of descriptor values increases approximately linearly with the number of carbon atoms. Equation VI has not reached yet semi-empirical methods, but can be used for fast and rough bond length prediction. Equation I obviously cannot be improved much in future. The advance of the analytical equations with respect to quantum-chemical calculations is in that the equations include quantities (descriptors) which are well defined and understandable to any chemist.

To reply on the question what is the chemical bond in a benzenoid hydrocarbon, we can say: the consequence of various phenomena, as follows from our analysis. At first point, it is the electron delocalization which defines the length of an aromatic bond, partially described by the  $\pi$  bond order. Then, here are integer parameters  $n$ ,  $m$ , and  $l$  which reflect the effects of the close (chemical) neighbourhood as delocalization, then steric interactions, electronegativity and ionic contribution to the chemical bond [56]. The effects of the far neighbourhood can be considered as small influences of the crystal packing due to various C – H $\cdots$ H, H $\cdots$ H, C( $\pi$ ) $\cdots$ C( $\pi$ ) i C( $\pi$ ) $\cdots$ H intermolecular interactions [18].

### CONCLUDING REMARKS

From the point of view of structural, theoretical and organic chemistry, the results shown by equations in Table 1 and Table 2 (I, II, VI) represent approximate or rather accurate solutions of the longly studied problem of bond length prediction for benzenoid hydrocarbons by empirical and simple theoretical methods. Chemometric approach in finding these equations clearly exhibits that the bond length in benzenoids depends on more quantities: the bond order, the chemical neighbourhood, and the effect of intermolecular interactions. Although some of the equations are as good as semi-empirical molecular orbital methods, this problem should be studied furthermore.

## Abstract

### BOND LENGTH PREDICTION FOR BENZENOID HYDROCARBONS

**Márcia M. C. Ferreira, Rudolf Kiralj, Universidade Estadual de Campinas, Campinas, SP, 13084-971, Brazil**

Carbon-carbon bond length empirical prediction has started already in 1930s, based on considerable set of crystallographic structures. Later on it was applied to aromatic compounds by means of linear regression methods and analytical curves. A historical review on the subject is described in this work. The results of the authors' research on bond length prediction for planar benzenoid hydrocarbons, as well as on application of chemometrics in this area, are presented also. The recent regression analysis by means of chemometric methods resulted in two important findings. First, benzenoid bond lengths in crystal are multivariate phenomenon closely related to bond orders, bond neighbourhood and crystal packing effects. Second, the established regression equations reach semi-empirical molecular orbital calculations, and could be further improved.

### LITERATURE

- [1] P. R. von Schleyer, H. Jiao, What is aromaticity?, *Pure Appl. Chem.*, **68** (1996) 209-218.
- [2] M. Glukhotsev, Aromaticity today: energetic and structural criteria, *J. Chem. Educ.*, **74** (1997) 132-136.
- [3] S. I. Kotelevskii, O. V. Prezhdo, Aromaticity indices revisited: refinement and application to certain five-membered ring heterocycles, *Tetrahedron*, **57** (2001) 5715-5729.
- [4] T. M. Krygowski, A. Ciesielski, C. W. Bird, A. Kotschy, Aromatic character of the benzene ring present in various topological environments in benzenoid hydrocarbons. Nonequivalence of indices of aromaticity, *J. Chem. Inf. Comput. Sci.*, **35** (1995) 203-210.
- [5] M. Cyrański, T. M. Krygowski, Separation of the energetic and geometric contributions to aromaticity. 3. Analysis of the aromatic character of benzene ring in their various derivatives, *J. Chem. Inf. Comput. Sci.*, **36** (1996) 1142-1145.
- [6] K. Woźniak, T. M. Krygowski, Crystallographic studies and physicochemical properties of  $\pi$ -electron systems. Part XVIII. Crystal and molecular structure of *N*-*n*-propyl-quinoxalinium 7,7',8,8'-tetracyano-*p*-quinodimethanide: estimation of charge at TCNQ by the HOSE model, *J. Mol. Struct.*, **193** (1989) 81-89.
- [7] T. M. Krygowski, A. Ciesielski, B. Świrski, P. Leszczyński, Variation of molecular geometry and aromatic character of chrysene and perylene in their EDA complexes. Refinement of X-ray crystal and molecular structure of chrysene and perylene, *Polish J. Chem.*, **68** (1994) 2097-2107.
- [8] A. R. Katritzky, M. Karelson, A. P. Wells, Aromaticity as a quantitative concept. 6. Aromaticity variation with molecular environment, *J. Org. Chem.*, **61** (1996) 1619-1623.
- [9] V. Bertolasi, P. Gilli, V. Ferretti, G. Gilli, Intermolecular N-H...O resonance. Heteroconjugated systems as hydrogen-bond-strengthening functional groups, *Acta Cryst.*, **B51** (1995) 1004-1015.
- [10] S. Tonmunpuean, V. Parasuk, S. Kokpol, QSAR study of antimalarial activities and artemisinin-heme binding properties obtained from docking calculations, *Quant. Struct.-Act. Relat.*, **19** (2000) 475-483.
- [11] R. Kiralj, Y. Takahata, M. M. C. Ferreira, QSAR of progestogens: use of a priori and computed molecular descriptors and molecular graphics, *QSAR Comb. Sci.*, **22** (2003) 430-448.
- [12] M. Tafazoli, A. Beaten, P. Geerlings, M. Kirschvolders, In vitro mutagenicity and genotoxicity study of a number of short-chain chlorinated hydrocarbons using the micronucleus test and the alkaline single cell gel electrophoresis technique (comet assay) in human lymphocytes – a structure-activity relationship (QSAR) analysis of the genotoxic and cytotoxic potential, *Mutagenesis*, **13** (1998) 115-126.
- [13] C. M. Marian, M. Gastreich, Quantitative structure-property relationships in boron nitrides: The N-15- and B-11 chemical shifts, *Solid State Nucl. Magn. Reson.*, **19** (2001) 29-44.

- [14] S. S. Liu, H. L. Liu, Z. N. Xia, C. Z. Cao, Z. L. Li, Molecular distance-edge vector  $\mu$ : and textension from alkanes to alcohols, *J. Chem. Inf. Comput. Sci.*, **39** (1999) 951-957.
- [15] A. J. Guttmann, M. Gege, I. Jensen, I. Gutman, The number of benzenoid hydrocarbons (in serbian), *Hem. Pregled*, **43** (2002) 105-108.
- [16] N. Trinajstić, Povratak povjerenja u teoriju valentnih struktura, *Kem. ind.*, **33** (1984) 499-514.
- [17] N. Trinajstić, *Chemical Graph Theory*, 2<sup>nd</sup> ed., CRC Press, Boca Raton, Florida, 1992.
- [18] R. Kiralj, M. M. C. Ferreira, Predicting bond lengths in planar benzenoid polycyclic aromatic hydrocarbons: a chemometric approach, *J. Chem. Inf. Comput. Sci.*, **42** (2002) 508-523.
- [19] F. K. Allen, O. Kennard, 3D search and research using the Cambridge Structural Database, *Chem. Design Autom. News*, **8** (1993) 131-137.
- [20] F. H. Allen, W. D. S. Motherwell, Applications of the Cambridge Structural Database in organic chemistry and crystal chemistry, *Acta Cryst.*, **B58** (2002) 407-422. Further information about the CSD database is available at <http://www.ccdc.cam.ac.uk>
- [21] R. Kiralj, B. Kojić-Prodić, M. Žinić, S. Alihodžić, N. Trinajstić, Bond length-bond order relationships and calculated geometries for some benzenoid aromatics, including phenanthridine. Structure of 5,6-dimethylphenanthridinium triflate, [N-(6-phenanthridinylmethyl)-aza-18-crown-6- $\kappa^5 O, O', O'', O''', O''''$ ](picrate- $\kappa^2 O, O'$ )potassium, and [N,N'-bis(6-phenanthridinyl- $\kappa N$ -methyl)-7,16-diaza-18-crown-6- $\kappa^4 O, O', O'', O'''$ ] sodium iodide dichloromethane solvate, *Acta Cryst.*, **B52** (1996) 823-837.
- [22] R. Kiralj, B. Kojić-Prodić, S. Nikolić, N. Trinajstić, Bond lengths and bond orders in benzenoid hydrocarbons and related systems: a comparison of valence bond and molecular orbital treatments, *J. Mol. Struct. (Theochem)*, **427** (1998) 25-37.
- [23] R. Kiralj, B. Kojić-Prodić, I. Piantanida, M. Žinić, Crystal and molecular structures of diazapyrenes and a study of  $\pi \dots \pi$  interactions, *Acta Cryst.*, **B55** (1999) 55-69
- [24] M. A. Sharaf, D. L. Illman, B. P. Kowalski, *Chemometrics*, J. Wiley&Sons, New York, 1986.
- [25] K. Beebe, R. Pell. M. B. Seasholtz, *Chemometrics: a practical guide*, John Wiley & Sons, New York, 1998.
- [26] M. M. C. Ferreira, A. M. Antunes, M. S. Melo, P. L. O. Volpe, *Quimiometria I: Calibração multivariada, um tutorial*, *Quím. Nova*, **22** (1999) 724-731.
- [27] M. M. C. Ferreira, *Multivariate QSAR*, *J. Braz. Chem. Soc.*, **13** (2002) 742-753.
- [28] M. M. C. Ferreira, C. A. Montanari, A. G. Gáudio, *Seleção de variáveis em QSAR*, *Quím. Nova*, **25** (2002) 439-448.
- [29] K. R. Beebe, B. P. Kowalski, An Introduction to multivariate calibration and analysis, *Anal. Chem.*, **59** (1987) 1007A-1017A.
- [30] M. M. C. Ferreira, Polycyclic aromatic hydrocarbons: a QSPR study, *Chemosphere*, **44** (2001) 125-146 (2001).
- [31] C. N. Alves, J. C. Pinheiro, A. J. Camargo, M. M. C. Ferreira, R. A. F. Romero, A. B. F. da Silva, A multiple linear regression and partial least squares study study of flavonoid compounds with anti-HIV activity, *J. Mol. Struct. (Theochem)*, **541** (2001) 81-88.
- [32] S. Subramanian, M. M. C. Ferreira, M. Tršić, A structure - activity relationship study of lapachol and some derivatives of 1,4-naphthoquinone against carcinosarcoma Walker 256, *Struct. Chem.*, **9** (1998) 47-57.
- [33] C. N. Alves, J. C. Pinheiro, A. J. Camargo, M. M. C. Ferreira, A. B. F. da Silva, A structure - activity relationship study of HEPT-analog compounds with anti-HIV activity, *J. Mol. Struct. (Theochem)*, **530** (2000) 39-47.
- [34] J. C. Pinheiro, M. M. C. Ferreira, O. A. S. Romero, Antimalarial activity of dihydroartemisinin derivatives against *P. falciparum* resistant to mefloquine: a quantum chemical and multivariate study, *J. Mol. Struct. (Theochem)*, **572** (2001) 35-44.
- [35] R. Kiralj, M. M. C. Ferreira, A priori molecular descriptors in QSAR: a case of HIV-1 inhibitors. I. The chemometric approach, *J. Mol. Graph. Mod.*, **21** (2003) 435-448.
- [36] R. Kiralj, M. M. C. Ferreira, A priori molecular descriptors in QSAR: a case of HIV-1 inhibitors. II. Molecular graphics and modeling, *J. Mol. Graph. Mod.*, **21** (2003) 499-515.
- [37] M. M. C. Ferreira, R. Kiralj: *Chemometric Methods in QSAR*. In: A. C. Montanari (Ed.): *Proceedings of the 1<sup>st</sup> Brazilian Symposium on Medicinal Chemistry* (portugues), in press.
- [38] R. Kiralj, M. M. C. Ferreira, Molecular Graphics-Structural and Molecular Graphics Descriptors in QSAR Study of 17- $\alpha$ -Acetoxypregesterones, *J. Braz. Chem. Soc.*, **14** (2003) 20-26.

- [39] L. Pauling, L. O. Brockway, J. Y. Beach, The dependence of interatomic distances on single bond – double bond resonance, *J. Am. Chem. Soc.*, **57** (1935) 2705-2709.
- [40] G. Lendvay, On the correlation of bond order and bond length, *J. Mol. Struct. (Theochem)*, **501-502** (2000) 389-393.
- [41] L. Pauling, The nature of the chemical bond, 2<sup>nd</sup> ed., Cornell University Press, Ithaca, New York, 1951.
- [42] L. Pauling, Bond numbers and bond lengths in tetrabenzo[*de, no, st, c<sub>1</sub>d<sub>1</sub>*]heptacene and other condensed aromatic hydrocarbons: a valence-bond treatment, *Acta Cryst.*, **B36** (1980) 1898-1901.
- [43] M. Randić, Graph theoretical derivation of Pauling bond orders, *Croat. Chem. Acta*, **47** (1987) 71-78.
- [44] L. Pauling, Atomic radii and interatomic distances in metals, *J. Am. Chem. Soc.*, **69** (1947) 542-553.
- [45] W. Gordy, Dependence of bond order and of bond energy upon bond length, *J. Chem. Phys.*, **15** (1947) 305-310.
- [46] R. Kiralj, M. M. C. Ferreira, On heteroaromaticity of nucleobases. Bond lengths as multidimensional phenomena, *J. Chem. Inf. Comput. Sci.*, **43** (2003) 787-809.
- [47] R. Goddard, M. W. Haenel, W. C. Herndon, C. Krüger, M. Zander, Crystallization of large planar polycyclic aromatic hydrocarbons: the molecular and crystal structures of hexabenzo[*bc, et, hi, kl, no, qr*]coronene and benzo[1,2,3-*bc*:4,5,6-*b'c'*]dicononene, *J. Amer. Chem. Soc.*, **117** (1995) 30-41.
- [48] S. Narita, T. Morikawa, T. Shibuya, Linear relationship between the bond lengths and the Pauling bond orders in fullerene molecules, *J. Mol. Struct. (Theochem)*, **532** (2000) 37-40.
- [49] T. Morikawa, S. Narita, T.-I. Shibuya, *J. Mol. Struct. (Theochem)*, **466** (1999) 137-143.
- [50] W. C. Herndon, Resonance theory. VI. Bond orders, *J. Amer. Chem. Soc.*, **96** (1974) 7605-7614.
- [51] W. C. Herndon, C. Párkányi,  $\pi$  Bond orders and bond lengths, *J. Chem. Educ.*, **53** (1976) 689-691.
- [52] B. P. Stoicheff, Variation of carbon-carbon bond lengths with environment as determined by spectroscopic studies of simple polyatomic molecules, *Tetrahedron*, **17** (1962) 135-145.
- [53] F. H. Allen, O. Kennard, D. G. Watson, L. Brammer, A. G. Orpen, Tables of bond lengths determined by X-ray and neutron diffraction. Part 1. Bond lengths in organic compounds, *J. Chem. Soc. Perkin Trans. 2*, (1987) S1-S19.
- [54] M. J. S. Dewar, G. J. Gleischer, G. J., Ground states of conjugated molecules. VII. Compounds containing nitrogen and oxygen, *J. Chem. Phys.*, **44** (1966) 759-773.
- [55] Titan v. 1.0.5., Wavefunction, Inc., Irvine, CA, 2000.
- [56] L. V. Vil'kov, V. S. Mastryukov, N. I. Sadova, Determination of the geometric structures of free molecules (english translation), Mir, Moscow, 1983, pp. 88-89.