Predicting Bond Lengths in Planar Benzenoid Polycyclic Aromatic Hydrocarbons: A Chemometric Approach†

R. Kiralj* and M. M. C. Ferreira

Laboratório de Quimiometria Teórica e Aplicada, Instituto de Química, Universidade Estadual de Campinas, Campinas, SP, 13083-970, Brazil

Received June 30, 2001

Two hundred and twenty-three aromatic carbon−carbon bond lengths in high precision crystal structures containing 22 planar condensed benzenoid polycyclic aromatic hydrocarbons (PB−PAHs) were related to the Pauling π-bond order, its analogue corrected to crystal packing effects, the number of hexagonal rings around the bond, and the numbers of carbons atoms around the bond at topological distance one and two. Principal Component Analysis (PCA) showed that the bond lengths in PB−PAHs are at least two-dimensional phenomenon, with well pronounced classification into 12 types of bonds, as confirmed with Hierachical Cluster Analysis (HCA). Consequently, Multiple Linear Regression (MLR) and Partial Least Squares (PLS) models were superior to univariate models, reducing the degeneration of the data set and improving the estimation of Julg's structural aromaticity index. The approximate regression models based on topological descriptors only were built for fast and easy prediction of bond lengths and bond orders in PB−PAHs.

1. INTRODUCTION

It is well-known that due to the size of atoms the chemical bond lengths d in organic molecules are usually 1−2 Å (with relative experimental error of structural determination ≈0.1%), and carbon−carbon d's are usually 1.2−1.6 Å.[1] For planar and nearly planar benzenoids (condensed benzenoid polycyclic aromatic hydrocarbons, PB−PAHs) the d range is 1.33−1.48 Å (10% variation).[2] The variations in d usually reflect differences in charge distribution around the bonds, and so bond lengths can be useful in structural, QSAR, or theoretical studies. Logical questions appear, like the following: Which properties, molecular descriptors (in fact, bond descriptors), define d? Is d one- or multidimensional phenomenon? How to predict d rapidly and easily, with enough accuracy?

Why to study carbon−carbon bonds? PB−PAHs, their derivatives, fragments, and heterocyclic analogues are widely abundant in synthetic and natural substances, especially in biological systems as the following: organic solvents, environmental carcinogens and mutagens, DNA-intercalators, constitutive parts of various drugs, nucleic acids and proteins, vitamins and coenzymes, etc. Intra- and intermolecular (hetero)aromatic-(hetero)aromatic stacking interactions of PB−PAHs and analogue fragments stabilize chemical and biological systems, especially in crystal phase.[3] Then, naturally, how to rationalize aromatic C−C, C−N, C−O and other d's if not starting with PB−PAHs as the simplest π-systems besides hexagonal graphite[4,5] and fullerenes?[6]

More papers studied d of PAHs or heterocyclic analogues, sometimes as heterogeneous sets (including planar and nonplanar PAHs, helicenes, even graphite and conjugated molecules as ethylene and butadiene) and in some cases without experimental accuracy, as unidimensional functions of the Pauling π-bond order p_P, $\log p_P$, other functions of p_P, other bond orders, or topological indices.[5−16] But today, as experimental techniques are more advanced than some two-three decades ago, usually new effects such as crystal packing effects, temperature, phase, quality of the crystal, and the measurement affect the final results of structural determinations. So what before was all inside the experimental errors, now seems to be out. In the absence of experimental known and unknown variables, ab initio study of bond length−bond order relationships seem to be more "clear" and rather unidimensional problem.[17] Then, which analysis, univariate or multivariate is preferred? Somebody interested in bond length predictions, especially a nonspecialist in graph theory or quantum chemistry, would ask is highly accurate prediction of d possible, or if not, then is there at least any fast and approximate prediction of d and p_P without complicated procedures.

Bond lengths and other structural variables derived from bond lengths are the structural criterion of aromaticity, one of the main aromaticity indices.[18−22] By other words, various indices on bond lengths equalization should point out how much some molecule, or ring, or other molecular fragment is aromatic, being closer or farther from benzene in the degree of π-electron delocalization. In general, variation in bond lengths of aromatic hydrocarbons (0.15 Å^2) is less than minimum variation between formal single and double bonds in antiaromatic hydrocarbons (0.2 Å[19]). Although being important, the structural aromaticity indices are not enough to have a complete idea on aromaticity of a molecule.[18,19,21]

As has already been said, the accuracy of the experimental data and the number of samples in the data set under the same or similar experimental conditions and database mining criteria may affect the quality and parameters of the

* Corresponding author phone: +55 19 3788 3102; fax.: +55 19 3788 3023; e-mail: rudolf@iqm.unicamp.br.

† Part of this work was presented at CC'97 Conferentia Chemometrica, August 21-23, 1997, Budapest, Hungary and at 24ª Reunião Anual da Sociedade Brasileira de Química, May 28-31, 2001, Poços de Caldas, MG, Brazil.

10.1021/ci010063g CCC: $22.00
Published on Web 03/09/2002

predictions. Cambridge Structural Database (CSD)[23] contains enormous number of crystal structures including PB−PAHs and other aromatic molecules in their crystals, in crystals of molecular complexes, in crystals where these molecules are solvates, clathrates, or ligands bound to metals. The number and the quality of PAHs structures in CSD seems to be a function of time. In his work in 1974 Herndon[11] used 13 PAHs with 100 unique (symmetrically independent) bonds, having average experimental estimated standard deviations on bond lengths $\sigma = 0.008$ Å, the correlation coefficient r between d and the Pauling π-bond order p_P being $r = 0.92$, and the average deviation of calculated from experimental d was $\Delta = 0.009$ Å. Herdon and Parkanyi[5] used practically the same set in 1976. Pauling[7] studied nine molecules with 82 unique bonds in 1980. Kiralj et al.[14] made a new search in CSD October 1995 Release finding 14 PB−PAHs with 124 unique bonds, $\sigma = 0.006$ Å, $r = 0.898$, $\Delta = 0.010$ Å. The next search[15] in CSD April 1996 Release resulted in 16 molecules with 147 bonds, $\sigma = 0.005$ Å, $r = 0.905$, $\Delta = 0.010$ Å. The last search[2] in CSD October 1998 Release gave 17 molecules and 153 bonds, $\sigma = 0.005$ Å, $r = 0.910$, $\Delta = 0.010$ Å. It is obvious that extensive mining in CSD and the most recent literature (for the newest structures which are not yet in current CSD Release since it is updated biannually) and update of d-p_P study is recommendable. Besides, from the statistical point of view, increase of data can reveal some new trends which have not been observed before. Here we extend previous studies[14−16] by updating the set of experimental d's for PB−PAHs, use multivariate versus univariate techniques to classify the bonds by Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA)[24] and to predict bond lengths using Multiple Linear Regression (MLR) and Partial Least Squares (PLS).[24,25] This study can be considered QSPR (Quantitative Structure−Property Relationship) since it relates experimental properties measured by X-ray or neutron diffraction methods to counted/calculated descriptors. The matter can be characterized also as structure correlation as variables in question are structural, electronic, or topological derived from 2D (chemical schemes) and 3D (experimental geometries) structures of PB−PAHs.

2. METHODOLOGY

a. Database Mining. The search for the best crystal structures (crystallographic $R < 0.07$, other criteria as defined before[14−16]) in CSD December 2000 Release[26] and in the most recent literature (1999−2001) was performed. The d values and their estimated standard deviations σ were averaged over maximum (gas phase) molecular symmetry and later on treated as unique bonds. For example, in the case of benzene whose molecule has C_i (crystallographic) symmetry and three different values for bond lengths, the bonds were averaged as $\langle d \rangle = (d_1 + d_2 + d_3)/3$ and their estimated standard deviations as $\sigma = (\sigma_1 + \sigma_2 + \sigma_3)/3$ to define a unique bond representing free benzene molecule with D_{6h} symmetry. This way of averaging, although not being statistically correct, treats all the unique bonds with the same weight, and from the point of view of chemical crystallography is justified. A few PB−PAHs with structures in CSD not suitable for the training/validation set were chosen for the prediction set.

b. Calculation of the Bond Descriptors for the Training/Validation Set and the Variable Selection. The bond descriptors were calculated without computer assistance as follows:

- p_P bond orders by empirical[5,7,27] or Randić method[5,27,28] for those molecules if not known from literature
- p_{cr} bond orders: p_P corrected to crystal packing effects (described below)
- n number: the number of C atoms around the bond (topological distance $t_D = 1$)
- m number: the number of hexagons around the bond
- l number: the number of C atoms around those atoms counted for n number ($t_D = 2$)
- k number: the number of C atoms around those atoms counted for l number ($t_D = 3$)
- analogous numbers i, j, υ at topological distances 4, 5, 6 from the considered bond
- m_{cr} numbers: m number corrected by adding its square, with coefficients found from parabolic fitting to d, p_P, and p_{cr}
- l_{cr} numbers: l number corrected by adding its square, cube, fourth to sixth powers, with coefficients found from sixth-order polynomial fitting to d, p_P, and p_{cr}

The estimation of p_P bond orders was based only on the first valence-bond approximation: only ground state (Kekulé) resonance structures were included with the same weight. Even in the case of formally single bonds, the first nonionic excited (Dewar) resonance structures were excluded, in contrary to Pauling.[7]

The n, m, l, k, i, j, υ numbers can be considered topological descriptors (indices). Stoicheff[29] showed that C−C bond lengths in organic molecules depend linearly on the number of carbon atoms bound to the bond under study. This integer variable (here index n) includes environmental effects as orbital hybridization, electron delocalization, steric interactions, electronegativity, and ionic contribution to bonding.[30] Applying this idea to formal C−C bonds[31] in organic crystals[1] the linear d-n relationship is shown to be fairly well established ($r = 0.967$, average deviation $\Delta = 0.009$ Å). This justifies estimation of n and other topological descriptors for PB−PAHs.

The maximum crystal packing effect on bond lengths (shortening or lengthening of a bond) is considered to be approximately 0.01−0.02 Å.[32] In the case of PAHs, where most of intermolecular contacts are of the type C−H...H, H...H, C(π)...C(π), C(π)...H, the maximum crystal packing effect is even smaller. Investigating this effect in the set of studied molecules by comparing the bonds which would be symmetrically equal in gas phase (structures with CSD REFCODEs: KEKULN10, BENZEN, PENCEN01, see the list of REFCODES in Appendix), using a method reported by Bürgi,[32] the following was concluded: shorter the bond, harder to deform it, so the crystal packing effect is proportional to p_P and can be considered being 0.001 Å as minimum and 0.007 Å as maximum. Calculated d_C's from d - p_P linear relationship showed that, when compared to experimental d's, p_P should be corrected in this way:

$$p_{cr} = p_P = p_P^0 \text{ if } |d - d_C| \leq 0.001 \text{ Å}$$

(no change in crystal)

$$p_{cr} = a_1 p_P + b_1 = p_P^- \text{ if } d - d_C > 0.001 \text{ Å}$$
(bond is shortened in crystal)

$$p_{cr} = a_2 p_P + b_2 = p_P^+ \text{ if } d_C - d > 0.001 \text{ Å}$$
(bond is lengthened in crystal)

where $a_1 = 1.041$, $a_2 = 0.959$, $b_1 = 0.007$, $b_2 = -0.007$.

Numerical (correlations with $r > 0.5$) and graphical studies of bond length−bond descriptor relationships were performed. Also, polynomial fits of n, m, l, k to d, p_P, and p_{cr} made clear that m_{cr} and l_{cr} could be used. Thus, the bond descriptors utilized in the further study were p_P, p_{cr}, n, m, l, m_{cr}, l_{cr}. Parabolic, logarithmic, and Pauling curve[7] (of the form $1.84x/(0.84x+1)$) fits of p_P and p_{cr} were also performed.

c. The Statistics of the Data Set Degeneration. Two bond lengths d_1 and d_2 from crystal structure determination are considered not to be significantly different (or "equal") at 0.99 probability level (normal distribution of the bond lengths in crystal is assumed) if

$$q = |d_1 - d_2|/[\sigma^2(d_1) + \sigma^2(d_2)]^{1/2} < 2.58$$

where $\sigma(d_1)$ and $\sigma(d_2)$ are estimated standard deviations of d_1 and d_2, respectively.[33] The degeneration statistics in this work is considered as the study of degenerated bond lengths, i.e. those which have the same value of one or more bond descriptors. This way, d-p_P, d-p_{cr}, d-(p_P, p_{cr}, n, m, l) relationships were studied. The data set was rearranged so that the bonds with equal values of considered descriptors come to the same group. Inside the groups, the number of comparisons (to see if the bonds are equal or not) was counted, and in some cases corresponding q values were estimated. Some other statistical parameters were calculated for the d-(p_P, p_{cr}, n, m, l) degeneration (see Results and discussion) using Matlab 5.4.[34]

d. Classification of the Bonds. Principal Component Analysis and Hierarchical Cluster Analysis were performed on training/validation autoscaled data sets. Pirouette 3.01[35] was employed. The identification of structural fragments around the bond, based on HCA dendogram and PC1−PC2 score plot, was performed. PCA for the prediction set was also carried out.

e. The Validation of the Regression Models. Unweighted linear regression (LR) models for calculation of d, based on bond orders p_P and p_{cr}, were performed using Matlab 5.4.[34] Unweighted MLR models for predicting d were build using the data sets (n, m, l), (n, m, l, m_{cr}, l_{cr}), (p_P, p_{cr}, n, m, l), and (p_P, p_{cr}, n, m, l, m_{cr}, l_{cr}). MLR was performed by Matlab 5.4.[34] Due to correlations between the bond descriptors, Principal Component Regression (PCR)[24,25] with all the Principal Components (PCs) was performed as equivalent to MLR. The PLS models were established in the very same way as MLR models. The MLR and PLS models were built to predict p_P and p_{cr} using (n, m, l) and (n, m, l, m_{cr}, l_{cr}) data sets as variables. Besides the standard validation parameter, average deviation Δ and average $q = \Delta/\sigma$ were used. The analysis was performing by Pirouette 3.01[35] on autoscaled data and leave-one-out crossvalidation was used for PLS and PCR models.

f. Calculation of the Bond Descriptors for the Prediction Set. Bond descriptors p_P, n, m, l were counted in the same way as for the training/validation set. The corrected Pauling π-bond order p_{cr} was calculated for both the training/validation and the prediction set, based on statistics observed in HCA/PCA analysis of the training/validation set. The procedure was performed in the following steps.

(1) Calculation of the number of unique bonds which should be unchanged (p_P^0), shortened (p_P^-), and lengthened (p_P^+) with respect to their calculated bond lengths from d-p_P linear relationship. The ratio of these bonds should be (p_P^0):(p_P^-):(p_P^+) = 10.3%:45.3%:44.4% as is for the 223 bonds.

(2) Initial bond distribution between (p_P^+) and (p_P^-). It is preferred to choose bonds with $n = 4$ for (p_P^+), especially if they occur closer to the molecular center or having greater neighborhood at topological distance $t_D = 2-4$. For (p_P^-) is the opposite: bonds with $n = 2$, far from molecular center, with more hydrogens at $t_D = 2-4$ are preferred. The rest, especially bonds with $n = 3$, should be distributed between (p_P^0), (p_P^-), and (p_P^+).

(3) Assigning the bonds for (p_P^0). All the rest are candidates for (p_P^0). At first, the bonds should be tested if satisfying the relationship $n + m + l = 15.574-11.253\ p_P$ found in the training/validation set for all (p_P^0) bonds ($r = 0.716$). Those bonds predicting $n + m + l$ with 0 or 1 as deviation from counted $n + m + l$ should remain for the further elimination step. Here, the bonds are rejected if their $n + m + l$ numbers are not main characteristics of the classes II−V, VIII, X−XI in PCA of the training/validation set (see Results and Discussion). If there are still more candidates than required, what could be possible due to high degeneration of the data set, bonds having $n + m + l$ different from 10 or 11 are rejected (the average $\langle n + m + l \rangle = 10.6$ for the training/validation set). Further elimination criteria are conditions $n = 3 \pm 1$, $m = 2 \pm 1$, $l = 5 \pm 1$ (the average values of the training/validation set are 3.2, 2.2, and 5.1, respectively). At the end of the numerical elimination, it is required that $n = 3$, $m = 2$, $l = 5$.

(4) If still there are more candidates than required, the distribution of the bonds for the three ways of calculating p_{cr} should be done optionally: to sign bonds with greater neighborhood as (p_P^+), more isolated bonds as (p_P^-), and some middle cases (frequently with $n = 3$) as (p_P^0). This way, some information on crystal packing effects was introduced into the prediction set. It is worth to note that the relations used here are not accurate, and so it can result in some loss of original information.

g. Predicting the Bond Lengths. The simplest and the best multivariate model was used to predict the bond lengths of the prediction set by Pirouette 3.01.[35] As an additional validation of the prediction, the predicted values were compared to the experimental data.

h. Julg's Aromaticity Index as Additional Validation Parameter. The multivariate model applied on the prediction set was compared with linear d-p_P and d-p_{cr} models by calculating the Julg's structural aromaticity index[18,21] A and average bond length $\langle d \rangle$ using expressions

$$A = 1 - 255\ [\Sigma/\langle d \rangle]^2$$

$$\Sigma = \Sigma_i\ [d_i - \langle d \rangle]^2/N$$

$$\sigma(\langle d \rangle) = [\Sigma_i\ \sigma^2(d_i)]^{1/2}/N$$

$$\sigma(\Sigma) = [\Sigma_i (d_i - \langle d \rangle)^2 \sigma^2(d_i)/\Sigma^2 + N \sigma^2(\langle d \rangle)]^{1/2}$$

$$\sigma(A) = 510 [\Sigma^2 \sigma^2(\Sigma)/\langle d \rangle^4 + \Sigma^4 \sigma^2(\langle d \rangle)/\langle d \rangle^6]^{1/2}$$

where σ's are standard deviations of $\langle d \rangle$, Σ and A, Σ is the standard deviation of the data set. A, $\langle d \rangle$ and their errors were calculated for the molecules, including all unique bonds with their multiplicities. Original, unaveraged bond lengths as well as bond lengths corrected to thermal motion were used for some molecules.

4. RESULTS AND DISCUSSION

The molecules under study are schematically presented in Figures 1 and 2. The results of PCA, HCA, and PLS are illustrated in Figures 3−6. Table 1 contains experimental, calculated, and predicted QSPR data for aromatic bonds in PB−PAHs. Bond length−bond descriptor correlations are presented in Table 2. The PCA results for the training/validation set are in Table 3. The regression models are compared in Table 4, and the experimental and calculated (by models 1, 2, 8) structural aromaticity indices are in Table 5.

a. Database Mining Results. The CSD and recent literature mining resulted as follows. There were found 22 high precision crystal structures containing 22 PB−PAHs (21 from CSD October 2000 release, one from the most recent literature, Figure 1), what comprised 223 symmetrically independent (unique) aromatic C−C bond lengths (Table 1) with average $\sigma = 0.005$ Å. This data set was treated as the training/validation set. Furthermore, five low precision crystal structures of PB−PAHs (Figure 2) with 86 symmetrically independent C−C bonds were used as the prediction set. This way, bond lengths from these five structures could be compared with predicted values, rather than making prediction for bond lengths without experimental values.

The training/validation set consists of 12 catacondensed and 10 pericondensed PB−PAHs, with variation in size from 1 to 15 hexagonal rings and 6−62 carbon−carbon bonds. The prediction set has one catacondensed and four pericondensed molecules, ranging from 4 to 11 rings and 21−53 bonds. In both sets, the number of ring is $n_r \approx 4 + 4\ n_b$ where n_b is the number of C−C bonds.

The bond lengths follow a normal distribution with almost 50% of the bonds in the range 1.405−1.435 Å and over 90% within 1.360−1.450 Å. The shortest bond is 1.331(2) and the longest 1.484(6) Å, corresponding to pure double and formally single bond,[1] respectively, which gives the maximum difference in lengths 0.153(6) Å.

b. Bond Length−Bond Descriptor Relationships for the Training/Validation Set. Table 2 shows that nonlinear regressions including bond orders p_P and p_{cr} are not significantly better than LR models, although they have been used in the literature.[7,13,17] The topological indices n, m, l, $m_{cr} = m + am^2$, $l_{cr} = l + al^2 + bl^3 + cl^4 + dl^5 + el^6$ are those used for the regression models. Among polynomial regressions, only those containing m_{cr} and l_{cr} satisfied the

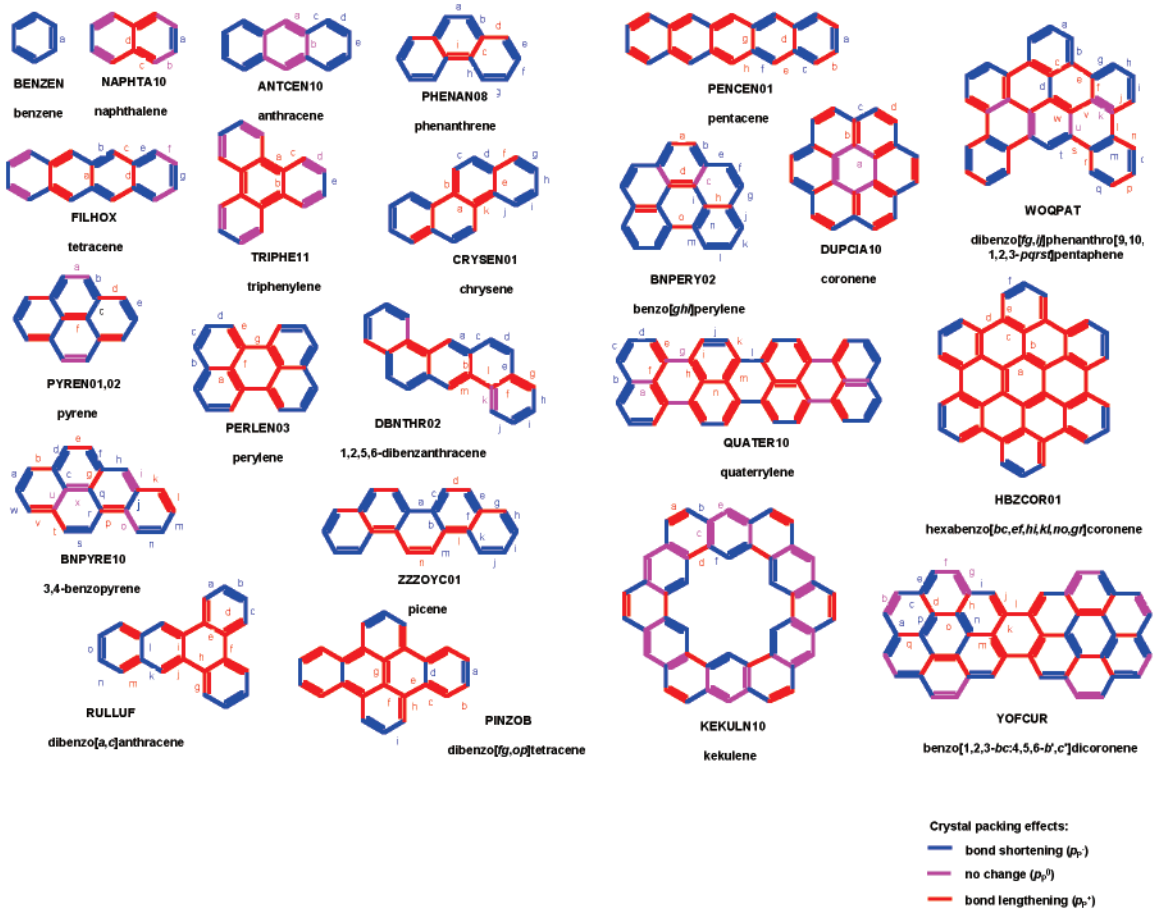

Figure 1. Molecules of the training/validation set represented by Kekulé structures, bond numeration, crystal packing effects, IUPAC names, and literature sources.

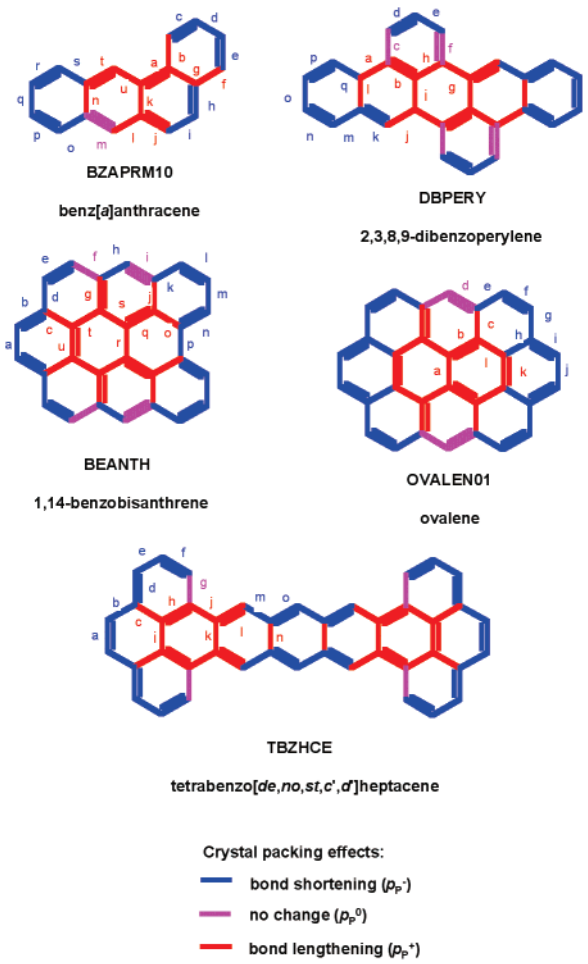

Figure 2. Molecules of the prediction set represented by Kekulé structures, bond numeration, crystal packing effects, IUPAC names, and literature sources.

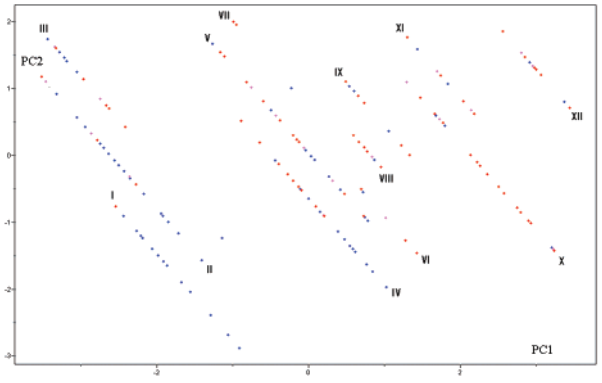

Figure 3. The PCA plot of the samples grouped into 12 groups and colored analogously to Figures 1 and 2.

condition $c_i/\sigma(c_i) \geq 2.58$ where c_i and $\sigma(c_i)$ are regression coefficient and its statistical error, respectively.

The correlation between d and topological indices decreases linearly in the order n-l-k (Table 2) and continues decreasing curvilinearly in order i-j-v ($r = 0.062$; -0.034; -0.067, respectively). This regular r-t_D relationship indicates that around $t_D = 5$ (approximately 5 Å, what corresponds to ovalene, Figure 2) there is no more influence of carbons atoms on the bond under consideration. By other words, the bond neighborhood of the size and shape of ovalene (Figure 2) ends after this limit.

There is a degeneration of data in terms of $d - p_P$ relationship as already noticed:[2,15] there are more d values (with differences beyond the experimental errors σ) with the same p_P values due to structural variations in the chemical bond, packing forces, unknown experimental errors, and other effects. This degeneration is more pronounced as the data set increases, and exists even in heterocyclic PB−PAHs analogues.[2,15] The previous study[2] on 17 PB−PAHs with 153 bonds revealed that 145 bonds are degenerated in 18 groups. There was a total of 682 comparisons between the bond lengths in the groups, q ranging up to around 12. About 30% of the comparisons had $q > 2.58$, implying the need of multivariate analysis. The most populate groups were those with $p_P = 0.300$ and 0.500 (9 bonds), $p_P = 0.400$ (10 bonds), and $p_P = 0.333$ (27 bonds). This high degeneration was expected as a consequence of the first valence-bond approximation, where only ground-state resonance structures with the same weight were used to calculate p_P. The set of 22 PB−PAHs in this work is even more degenerated, having 216 degenerated bonds spread in 40 groups. There are 1040 comparisons, and around 30% of them are significantly different, $q > 2.58$. The most populated groups are those with $p_P = 0.200$ (10 bonds), $p_P = 0.300$ (12 bonds), $p_P = 0.333$ (28 bonds), $p_P = 0.400$ (19 bonds), $p_P = 0.500$ (17 bonds), and $p_P = 0.667$ (11 bonds). These six groups can be easily observed if the groups average $\langle d_{gr} \rangle$ (including the seven one-membered groups) is plotted vs their population f_{gr} (ranging $\langle d_{gr} \rangle = 1.37-1.45$ Å) in normal distribution with maximum at 1.417 Å ($p_P = 0.333$). The bond lengths variability inside the groups is observed easily when studying the standard deviation of the group σ_{gr}. The σ_{gr} values reach maximum value of 0.034 Å. The σ_{gr} vs f_{gr} plot shows that the group standard deviations are mainly in the range 0.005−0.018 Å, with this maximum at 1.417 Å ($\sigma_{gr} = 0.012$ Å). On the other side, the plot $\langle d_{gr} \rangle$ vs σ_{gr} reveals three main regions with maximum in σ_{gr}. One is related to highly localized double bonds (ranging 1.37−1.35 Å, the highest peak of 0.017 Å is at 1.347 Å), the other is placed around the benzene value 1.390(9) Å (ranging 1.37−1.40 Å, having the peak of 0.030 Å at 1.388 Å), and the third is after the graphite value 1.422(1) Å (ranging 1.42−1.47, with the peak of 0.034 Å at 1.430 Å). Besides that, the maximum difference between bond lengths in a group Δ_{gr} reaches 0.078 Å. Although 65% of comparisons belong to the six mostly populated groups, Δ_{gr} is high also for low populated groups.

Multivariate models should decrease the degeneration of the data. Introducing crystal packing effects, p_{cr} becomes a two-dimensional function as it depends on both p_P and corrections for the packing effects. The d-p_{cr} relationship is characterized by 188 degenerated bonds in 48 groups. The most populated classes are those with $p_{cr} = 0.353$ (13 bonds) and $p_{cr} = 0.313$ (12 bonds), indicating that the highest degeneration is still at the graphitic bond region. There are 447 comparisons altogether.

When considering the degeneration with respect to the set (p_P, p_{cr}, n, m, l), 111 bonds in 44 groups are found as degenerated. There are only 98 comparisons, which means 9.4% of the initial number of 1040. Maximum $q = 6.36$, and there are only 2.2% of 1040 comparisons with $q > 2.58$.

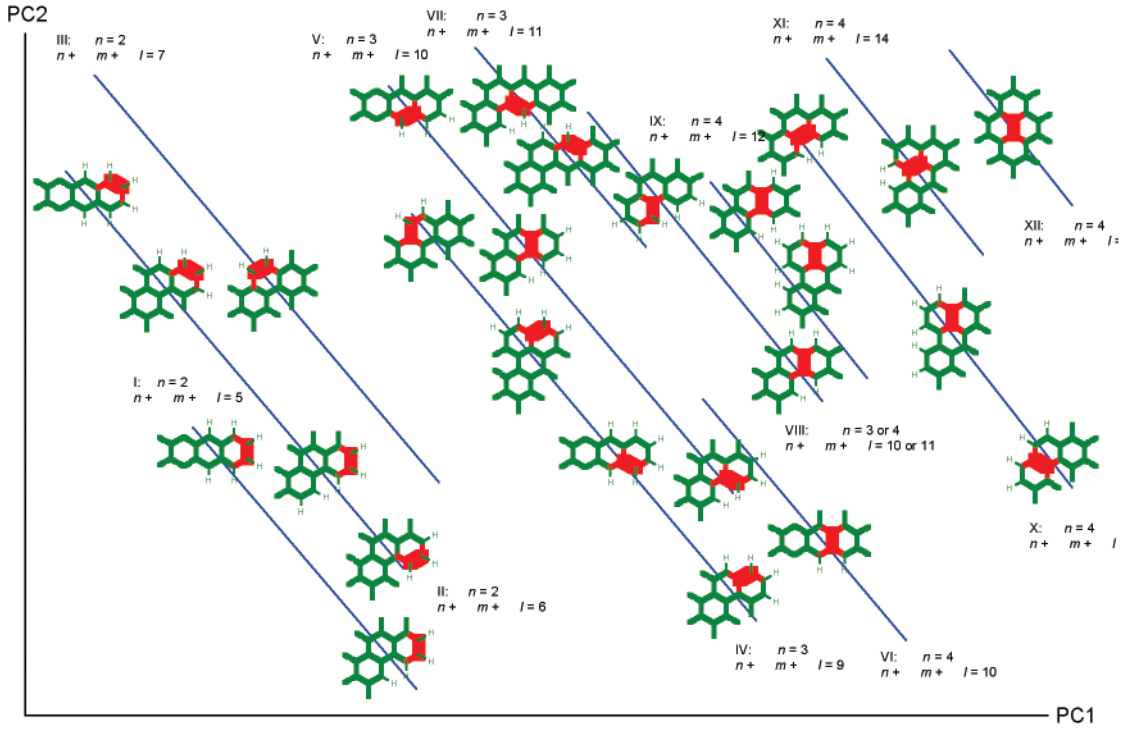

Figure 4. The PCA classes of C−C bonds. The dominant common fragments and bond descriptors are shown.

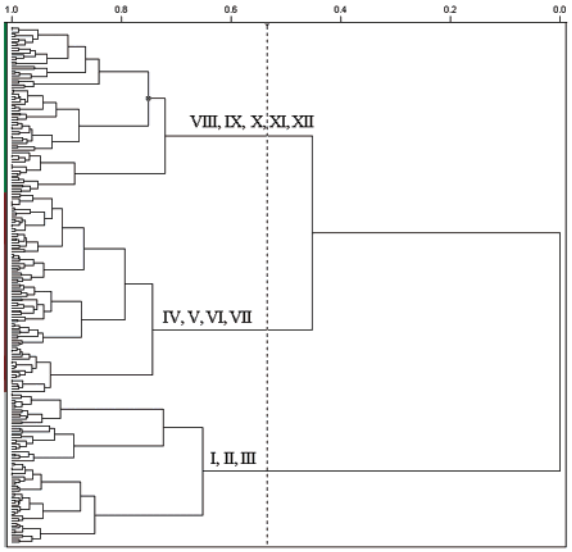

Figure 5. The HCA dendogram showing the distribution of classes I−XII.

This nicely illustrates how multivariate analysis is a powerful tool even when experimental errors are included into the study.

c. PCA and HCA Study of the Training/Validation Data Set. Table 3 presents the results of PCA on bond descriptors (p_P, p_{cr}, n, m, l). It is clear that the aromatic carbon−carbon bond length in PB−PAHs is at least two-dimensional phenomenon (96.5% of the total variance explained by the first two PCs), which is in accordance with the early observations by Dewar and Gleischer[12] on aromatic bond lengths. PCA on the prediction data set (using the same descriptors) confirms this observation. Even Herndon and Párkányi[5] realized that p_P bond order described at most 85% variation in bond lengths. On the other side, expected high correlations between p_P and p_{cr} and moderate correlations between n, m, l (it can be shown that for a given n only definite values of m are possible, and the same is valid with respect to l for a given m) explain the successful compression of the data. Including m_{cr} and l_{cr} into the data set does not increase information in PC1 and PC2 but rather makes less clear the relationships among the samples and the variables. Furthermore, all the bond descriptors are important for PC1 and PC2 (p_P and p_{cr} have high negative values at PC1 while n, m, l have positive).

There are 12 C−C bond classes (lines I−XII in Figure 3), in general well characterized by number $n + m + l$ which increases as PC1 increases. For classes I−III $n = 2$, for IV and V $n = 3$, for IX-XII $n = 4$, and VI−VIII disturb the regularity. The bonds are arranged in a strict two-dimensional pattern, in parallel lines and inside the lines. The dominant structural fragments in Figure 4 help to visualize the both trends. The neighborhood of particular C−C bond becomes denser (less hydrogens and empty space, more carbons around) as PC1 increases and PC2 decreases (Figure 4). This behavior has some analogy to that one in the above discussion on bond length−bond descriptor relationship. The bigger the bond neighborhood, the longer the bond and the closer to the graphitic bond. Having many carbons around, the bond becomes far from hydrogens, and it seems like there are graphitic fragments inside the molecules, which can be concluded when bond lengths are considered inside coronene and its derivatives. The bonds with one or two hydrogens seem to undergo more single bond−double bond alternation. The bonds which are formally single have no Kekulé structure in which they would be double. That is why the 223 bonds show complex behavior. It is interesting to note that the training/validation set contains bonds in all theoretically possible classes appearing in PB−PAHs (12 classes

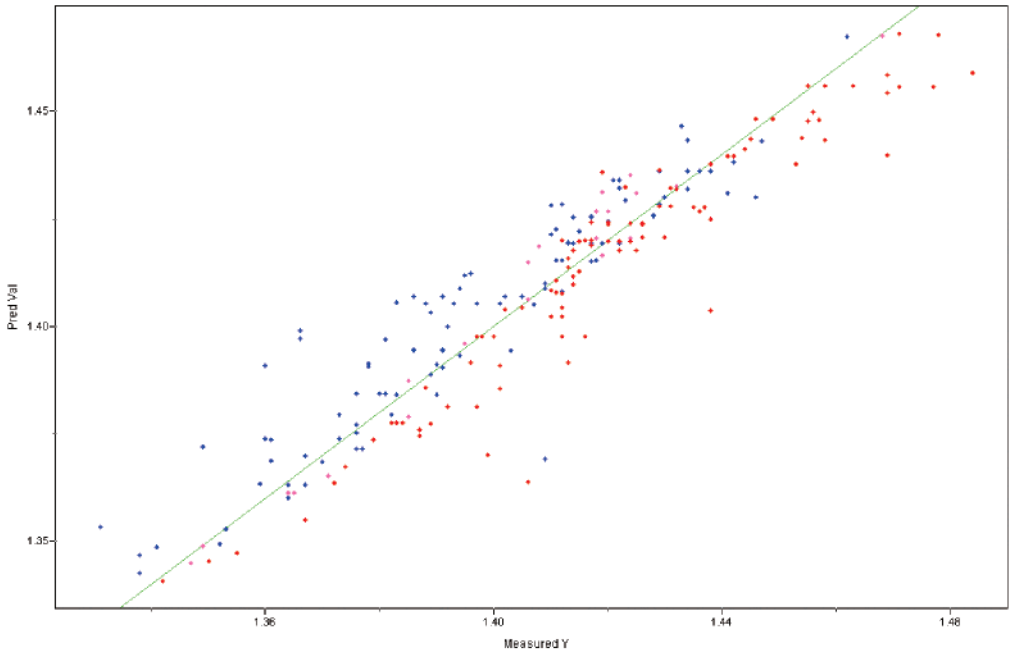

Figure 6. The PLS plot for model 8, colored in the same way as in Figures 1−3.

defined by *n*, *m*, *l* numbers). The same is with the prediction set. Observing the results of PCA in Table 3 and applying to Figure 4, the regularity between the lines and inside each one can be rationalized. The bonds are affected more by topological descriptors and not much by bond orders if PC1 increases (both *n* and $n + m + l$ increase, the bonds are more surrounded by carbon atoms). The opposite is when PC1 decreases (terminal bonds depending practically on bond orders only). When PC2 increases, the bonds are affected by bond orders and topological descriptor *l*, so that the bond gets shorter and has less carbons around it at $t_D = 3$. By other words, molecular fragment around the bond varies in its shape becoming more symmetric and compact and less branched if PC2 increases. In this sense, we can outline that PC1 is more connected to bond length and PC2 with the shape of molecular neighborhood (what is equivalent to the bond position in a molecule).

Colored Figures 1 and 2 show the distribution of the bonds with p_P^0, p_P^+, and p_P^-, i.e. not changed, lengthened, and shortened bonds in crystal with respect to the predicted values based on p_P only. A general conclusion can be outlined that most of the inner bonds are lengthened and the outer are shortened. Such a discrimination in C−C aromatic bond lengths in bond length−bond order relationship studies has never been observed before. This phenomenon could be partially originated from molecular structure effects (described by *n*, *m*, *l* bond descriptors) and crystal packing effect (see the discussion on PLS models). Figure 3 demonstrates the same color classification of the bonds. The lengthened bonds are more frequent as PC1 and PC2 increase, the opposite is with the shortened bonds, while bonds without change are almost uniformly distributed in the PC1−PC2 space. Figure 4 also helps to understand this trend.

Bond variables form two clusters, (p_P, p_{cr}) and (*n*, *m*, *l*), in HCA. In general, C−C bonds are grouped in classes in a similar way as in PCA (Figure 5).

d. Univariate Regressions. The prediction of *d* (training/validation set) using bond orders reaches $r = 0.94$ when p_{cr} is used in log *x* form. The use of squares of p_P or p_{cr} (parabolic regression) slightly improves the model. The most reasonable choice are linear models, among which model 2 should be the preferred one (Table 4).

The regression equation $d/\text{Å} = 1.467(2) - 0.147(5)\, p_P$ is not different than those in the previous works.[2,14,15] The coefficient *b*[2,14,15] is greater than by Herndon[11] and Herndon and Párkányi[5] since they included conjugated species such as ethylene and butadiene and even graphite. The equation $\langle d/\text{Å}\rangle = 1.470-0.153\, p_P$ for the 47 groups of degenerated data (see the above discussion on the degeneration of d-p_P relationship) is much less multidimensional ($r = 0.969$), due to loss of information when averaging the *d*'s. This again confirms that *d* is not a univariate problem.

The analogous regression equation $d/\text{Å} = 1.470(2) - 0.151(4)\, p_{cr}$ uses p_{cr} values based on analytical corrections of p_P as described in the methodology section. All regression parameters (Table 4) show that it is better to use p_{cr} instead of p_P. Besides, when p_{cr} was calculated by procedure as recommended for the prediction set, some information was lost so the resulting d-p_{cr} correlation was not significantly better than d-p_P ($r = -0.899$). This confirms that, although complicated, the procedures to calculate p_{cr} for the training/validation and the prediction sets were not based on wrong assumptions.

e. Multivariate Regression Models. The models for prediction of *d* (training/validation set) reaches $r = 0.96$ (models 1−10, Table 4). The best and the most parsimoneous model to propose is model 8 presented in Figure 6. The calculated and predicted *d* values for this model are in Table 1. It is obvious that m_{cr} and l_{cr} do not bring new information and that without the bond orders the models get worse. MLR models with many variables have $c_i/\sigma(c_i) > 2.58$, although numerically it can be avoided by performing the equivalent PCR. PLS with just a few PCs (two or three) would be the simplest and the best model, containing compressed and the most significant information, and high correlations between the bond descriptors are eliminated. The multivariate predic-

Table 1. Carbon−Carbon Bond Descriptors[a] for Planar Benzenoid PAHs

no.	molecule	bond	p_P	p_{cr}	n	m	l	d_{exp}/Å	d_{calc}/Å	σ/Å
1	benzene	a	0.500	0.527	2	1	2	1.390	1.384	0.009
2	naphthalene	a	0.333	0.353	2	1	2	1.407	1.405	0.002
3		b	0.667	0.667	2	1	3	1.371	1.365	0.002
4		c	0.333	0.313	3	2	4	1.422	1.418	0.002
5		d	0.333	0.313	4	2	4	1.420	1.424	0.002
6	anthracene	a	0.500	0.500	3	2	5	1.395	1.396	0.003
7		b	0.250	0.250	4	2	4	1.432	1.433	0.003
8		c	0.250	0.267	3	2	4	1.428	1.426	0.003
9		d	0.750	0.787	2	1	3	1.353	1.353	0.004
10		e	0.250	0.267	2	1	2	1.418	1.416	0.005
11	phenanthrene	a	0.800	0.839	2	1	4	1.338	1.347	0.005
12		b	0.200	0.215	3	2	4	1.422	1.432	0.007
13		c	0.400	0.377	4	2	5	1.413	1.416	0.007
14		d	0.400	0.377	3	2	4	1.414	1.410	0.004
15		e	0.600	0.631	2	1	3	1.349	1.372	0.008
16		f	0.400	0.423	2	1	2	1.381	1.397	0.009
17		g	0.600	0.631	2	1	3	1.376	1.372	0.004
18		h	0.400	0.423	3	2	5	1.391	1.407	0.007
19		i	0.200	0.185	4	3	6	1.454	1.444	0.006
20	tetracene	a	0.200	0.185	4	2	4	1.442	1.440	0.002
21		b	0.400	0.423	3	2	5	1.405	1.407	0.002
22		c	0.600	0.569	3	2	5	1.388	1.386	0.002
23		d	0.200	0.185	4	2	4	1.441	1.440	0.002
24		e	0.200	0.215	3	2	4	1.434	1.432	0.002
25		f	0.800	0.800	2	1	3	1.349	1.349	0.002
26		g	0.200	0.215	2	1	2	1.415	1.422	0.002
27	triphenylene	a	0.111	0.100	4	3	6	1.469	1.455	0.008
28		b	0.444	0.419	4	2	6	1.411	1.411	0.009
29		c	0.444	0.419	3	2	5	1.405	1.405	0.009
30		d	0.556	0.556	2	1	3	1.385	1.379	0.010
31		e	0.444	0.469	2	1	2	1.390	1.391	0.011
32	chrysene	a	0.500	0.473	4	2	6	1.402	1.404	0.002
33		b	0.250	0.233	3	2	5	1.437	1.428	0.002
34		c	0.750	0.787	2	1	4	1.331	1.353	0.002
35		d	0.250	0.267	3	2	4	1.417	1.426	0.002
36		e	0.375	0.353	4	2	5	1.417	1.419	0.002
37		f	0.375	0.353	3	2	4	1.415	1.413	0.002
38		g	0.625	0.657	2	1	3	1.361	1.369	0.002
39		h	0.375	0.397	2	1	2	1.392	1.400	0.002
40		i	0.625	0.657	2	1	3	1.370	1.369	0.002
41		j	0.375	0.397	3	2	5	1.409	1.410	0.002
42		k	0.250	0.233	4	3	6	1.453	1.438	0.002
43	pyrene	a	0.833	0.833	2	1	4	1.347	1.345	0.004
44		b	0.167	0.181	3	2	4	1.429	1.436	0.004
45		c	0.333	0.353	4	3	6	1.417	1.425	0.003
46		d	0.500	0.473	3	2	4	1.400	1.398	0.004
47		e	0.500	0.527	2	1	3	1.380	1.384	0.005
48		f	0.333	0.313	4	4	8	1.423	1.432	0.003
49	perylene	a	0.333	0.313	4	2	6	1.426	1.424	0.002
50		b	0.333	0.353	3	2	4	1.411	1.415	0.002
51		c	0.667	0.701	2	1	3	1.359	1.363	0.003
52		d	0.333	0.353	2	1	3	1.393	1.405	0.002
53		e	0.667	0.633	3	2	5	1.384	1.378	0.002
54		f	0.333	0.313	4	3	7	1.429	1.428	0.002
55		g	0.000	−0.007	4	3	6	1.471	1.468	0.002
56	1,2,5,6-dibenz-	a	0.500	0.527	3	2	5	1.391	1.394	0.002
57	anthracene	b	0.333	0.313	4	2	5	1.426	1.424	0.002
58		c	0.167	0.181	3	2	4	1.438	1.436	0.002
59		d	0.833	0.874	2	1	4	1.338	1.343	0.002
60		e	0.167	0.181	3	2	4	1.436	1.436	0.002
61		f	0.417	0.393	4	2	5	1.413	1.414	0.002
62		g	0.417	0.393	3	2	4	1.412	1.408	0.002
63		h	0.583	0.614	2	1	3	1.360	1.378	0.002
64		i	0.417	0.441	2	1	2	1.403	1.395	0.002
65		j	0.583	0.614	2	1	3	1.373	1.374	0.002
66		k	0.417	0.417	3	2	5	1.406	1.406	0.002
67		l	0.167	0.153	4	3	6	1.455	1.448	0.002
68		m	0.500	0.473	3	2	5	1.397	1.398	0.002
69	picene	a	0.308	0.327	4	3	6	1.429	1.429	0.007
70		b	0.462	0.488	4	2	6	1.388	1.405	0.009
71		c	0.231	0.247	3	2	5	1.412	1.428	0.009
72		d	0.769	0.731	2	1	4	1.367	1.355	0.008
73		e	0.231	0.247	3	2	4	1.410	1.428	0.010
74		f	0.385	0.362	4	2	5	1.414	1.418	0.009

Table 1 (Continued)

no.	molecule	bond	p_P	p_{cr}	n	m	l	d_{exp}/Å	d_{calc}/Å	σ/Å
75		g	0.385	0.362	3	2	4	1.414	1.412	0.008
76		h	0.615	0.647	2	1	3	1.367	1.370	0.009
77		i	0.385	0.408	2	1	2	1.366	1.399	0.010
78		j	0.615	0.647	2	1	3	1.409	1.370	0.009
79		k	0.385	0.408	3	2	5	1.394	1.409	0.010
80		l	0.231	0.215	4	3	6	1.469	1.440	0.008
81		m	0.308	0.289	3	2	5	1.430	1.421	0.010
82		n	0.692	0.657	2	1	4	1.406	1.365	0.008
83	3,4-benzopyrene	a	0.444	0.469	2	1	3	1.378	1.391	0.013
84		b	0.556	0.527	3	2	4	1.401	1.391	0.012
85		c	0.333	0.353	4	3	6	1.414	1.425	0.012
86		d	0.111	0.122	3	2	4	1.434	1.443	0.012
87		e	0.889	0.846	2	1	4	1.342	1.341	0.013
88		f	0.111	0.122	3	2	4	1.447	1.443	0.012
89		g	0.222	0.206	4	3	6	1.444	1.441	0.011
90		h	0.667	0.701	3	2	5	1.361	1.373	0.012
91		i	0.333	0.333	3	2	5	1.419	1.417	0.012
92		j	0.333	0.353	4	2	5	1.410	1.421	0.011
93		k	0.333	0.313	3	2	4	1.425	1.418	0.012
94		l	0.667	0.633	2	1	3	1.374	1.368	0.014
95		m	0.333	0.353	2	1	2	1.397	1.405	0.014
96		n	0.667	0.701	2	1	3	1.364	1.363	0.013
97		o	0.333	0.333	3	2	5	1.419	1.417	0.012
98		p	0.333	0.313	4	3	6	1.435	1.428	0.012
99		q	0.444	0.469	4	3	7	1.395	1.412	0.011
100		r	0.222	0.238	3	2	5	1.423	1.429	0.011
101		s	0.778	0.817	2	1	4	1.352	1.349	0.012
102		t	0.222	0.206	3	2	4	1.441	1.431	0.012
103		u	0.333	0.333	4	3	6	1.418	1.427	0.011
104		v	0.444	0.419	3	2	4	1.412	1.404	0.012
105		w	0.556	0.585	2	1	3	1.376	1.377	0.014
106		x	0.333	0.333	4	4	8	1.419	1.431	0.011
107	pentacene	a	0.167	0.181	2	1	2	1.428	1.426	0.005
108		b	0.833	0.792	2	1	3	1.355	1.348	0.006
109		c	0.167	0.181	3	2	4	1.434	1.436	0.005
110		d	0.167	0.153	4	2	4	1.445	1.444	0.005
111		e	0.667	0.657	3	2	5	1.387	1.376	0.005
112		f	0.333	0.353	3	2	5	1.412	1.415	0.005
113		g	0.167	0.153	4	2	4	1.458	1.444	0.005
114		h	0.500	0.473	3	2	5	1.412	1.398	0.005
115	dibenzo[*a*,*c*]-	a	0.538	0.567	2	1	3	1.382	1.379	0.001
116	anthracene	b	0.462	0.488	2	1	2	1.389	1.389	0.001
117		c	0.538	0.567	2	1	3	1.373	1.379	0.001
118		d	0.462	0.436	3	2	5	1.412	1.402	0.001
119		e	0.462	0.436	4	2	6	1.410	1.408	0.001
120		f	0.077	0.067	4	3	6	1.469	1.459	0.001
121		g	0.462	0.436	3	2	5	1.410	1.402	0.001
122		h	0.077	0.067	4	3	6	1.469	1.459	0.001
123		i	0.308	0.289	4	2	6	1.436	1.427	0.001
124		j	0.692	0.657	3	2	6	1.387	1.375	0.001
125		k	0.385	0.408	3	2	5	1.409	1.409	0.001
126		l	0.308	0.327	4	2	4	1.420	1.424	0.001
127		m	0.308	0.289	3	2	4	1.426	1.421	0.001
128		n	0.692	0.727	2	1	3	1.364	1.360	0.001
129		o	0.308	0.327	2	1	2	1.412	1.408	0.001
130	dibenzo[*fg*,*op*]-	a	0.450	0.475	2	1	2	1.391	1.390	0.005
131	tetracene	b	0.550	0.521	2	1	3	1.397	1.382	0.005
132		c	0.450	0.425	3	2	5	1.438	1.404	0.005
133		d	0.450	0.475	4	2	6	1.386	1.407	0.005
134		e	0.100	0.089	4	3	6	1.477	1.456	0.005
135		f	0.400	0.377	4	3	7	1.416	1.420	0.005
136		g	0.200	0.185	4	4	8	1.457	1.448	0.005
137		h	0.500	0.473	3	2	5	1.416	1.398	0.005
138		i	0.500	0.527	2	1	3	1.383	1.384	0.005
139	benzo[*ghi*]-	a	0.643	0.610	2	1	4	1.399	1.371	0.008
140	perylene	b	0.357	0.378	3	2	4	1.396	1.412	0.007
141		c	0.429	0.429	4	3	6	1.406	1.415	0.007
142		d	0.286	0.268	4	4	8	1.438	1.438	0.006
143		e	0.214	0.230	3	2	4	1.446	1.430	0.008
144		f	0.786	0.825	2	1	4	1.341	1.348	0.008
145		g	0.214	0.230	3	2	4	1.430	1.430	0.007
146		h	0.357	0.336	4	3	6	1.438	1.425	0.006
147		i	0.286	0.304	4	4	8	1.419	1.436	0.006
148		j	0.429	0.453	3	2	4	1.389	1.403	0.007

Table 1 (Continued)

no.	molecule	bond	p_P	p_{cr}	n	m	l	d_{exp}/Å	d_{calc}/Å	σ/Å
149		k	0.571	0.601	2	1	3	1.376	1.375	0.008
150		l	0.429	0.453	2	1	3	1.394	1.393	0.007
151		m	0.571	0.571	3	2	5	1.385	1.387	0.007
152		n	0.357	0.378	4	3	7	1.411	1.423	0.006
153		o	0.071	0.061	4	3	6	1.484	1.460	0.006
154	coronene	a	0.300	0.300	4	4	8	1.424	1.435	0.005
155		b	0.400	0.377	4	3	6	1.420	1.420	0.005
156		c	0.300	0.319	3	2	4	1.414	1.419	0.005
157		d	0.700	0.665	2	1	4	1.372	1.364	0.005
158	dibenzo[*fg,ij*]-	a	0.500	0.527	2	1	3	1.381	1.384	0.003
159	phenanthro[9,10,	b	0.500	0.527	3	2	5	1.386	1.394	0.003
160	1,2,3-*pqrst*]-	c	0.400	0.377	4	3	7	1.424	1.420	0.002
161	pentaphene	d	0.200	0.215	4	4	8	1.433	1.446	0.003
162		e	0.100	0.089	4	3	6	1.463	1.456	0.002
163		f	0.500	0.473	4	3	7	1.411	1.408	0.003
164		g	0.400	0.423	3	2	5	1.402	1.407	0.003
165		h	0.600	0.631	2	1	3	1.377	1.372	0.003
166		i	0.400	0.423	2	1	3	1.366	1.397	0.003
167		j	0.600	0.569	3	2	5	1.401	1.386	0.003
168		k	0.300	0.300	4	3	7	1.425	1.431	0.002
169		l	0.100	0.089	4	3	6	1.455	1.456	0.003
170		m	0.350	0.371	4	2	6	1.413	1.419	0.002
171		n	0.550	0.521	3	2	5	1.413	1.392	0.003
172		o	0.450	0.475	2	1	3	1.360	1.391	0.003
173		p	0.550	0.521	2	1	2	1.392	1.381	0.003
174		q	0.450	0.475	2	1	3	1.378	1.391	0.003
175		r	0.550	0.521	3	2	5	1.396	1.392	0.003
176		s	0.100	0.089	4	3	6	1.471	1.456	0.002
177		t	0.400	0.527	3	2	6	1.391	1.395	0.002
178		u	0.400	0.400	4	3	7	1.408	1.419	0.003
179		v	0.200	0.185	4	4	8	1.449	1.448	0.002
180		w	0.400	0.377	4	4	8	1.424	1.424	0.002
181	quaterrylene	a	0.333	0.333	4	3	6	1.420	1.427	0.004
182		b	0.333	0.353	3	2	4	1.417	1.415	0.004
183		c	0.667	0.701	2	1	3	1.367	1.363	0.004
184		d	0.333	0.353	2	1	3	1.401	1.405	0.004
185		e	0.667	0.633	3	2	5	1.382	1.378	0.004
186		f	0.333	0.313	4	3	7	1.431	1.428	0.004
187		g	0.000	0.000	4	3	6	1.468	1.468	0.004
188		h	0.333	0.313	4	3	7	1.431	1.428	0.004
189		i	0.667	0.633	3	2	5	1.383	1.378	0.004
190		j	0.333	0.353	2	1	4	1.383	1.405	0.004
191		k	0.667	0.633	3	2	5	1.389	1.378	0.004
192		l	0.000	0.007	4	3	6	1.462	1.467	0.004
193		m	0.333	0.313	4	3	7	1.429	1.428	0.004
194		n	0.333	0.313	4	4	8	1.431	1.432	0.004
195	hexabenzo[*bc,ef,*	a	0.400	0.377	4	4	8	1.417	1.424	0.002
196	*hi,kl,no,qr*]-	b	0.200	0.185	4	4	8	1.446	1.448	0.002
197	coronene	c	0.400	0.377	4	3	7	1.417	1.420	0.002
198		d	0.100	0.089	4	3	6	1.458	1.456	0.002
199		e	0.500	0.473	3	2	5	1.398	1.398	0.002
200		f	0.500	0.527	2	1	3	1.376	1.384	0.002
201	kekulene	a	0.850	0.809	2	1	4	1.350	1.346	0.002
202		b	0.150	0.163	3	2	4	1.442	1.438	0.002
203		c	0.350	0.350	4	2	5	1.418	1.420	0.002
204		d	0.150	0.137	4	3	6	1.456	1.450	0.002
205		e	0.500	0.500	3	2	5	1.395	1.396	0.002
206		f	0.500	0.527	3	2	6	1.386	1.395	0.002
207	benzo[1,2,3-	a	0.300	0.319	3	2	4	1.417	1.419	0.002
208	*bc*:4,5,6-	b	0.700	0.700	2	1	4	1.364	1.361	0.002
209	*b′,c′*]dicoronene	c	0.300	0.319	3	2	4	1.422	1.419	0.002
210		d	0.400	0.377	4	3	6	1.415	1.420	0.002
211		e	0.300	0.319	3	2	4	1.419	1.419	0.002
212		f	0.700	0.700	2	1	4	1.365	1.361	0.002
213		g	0.300	0.300	3	2	4	1.424	1.421	0.002
214		h	0.400	0.377	4	3	6	1.412	1.420	0.002
215		i	0.300	0.319	3	2	5	1.413	1.420	0.002
216		j	0.700	0.665	3	2	6	1.379	1.374	0.002
217		k	0.300	0.281	4	3	7	1.432	1.432	0.002
218		l	0.000	−0.007	4	3	6	1.478	1.468	0.002
219		m	0.400	0.377	4	4	8	1.420	1.424	0.002
220		n	0.300	0.319	4	4	8	1.421	1.434	0.002
221		o	0.300	0.281	4	4	8	1.429	1.436	0.002
222		p	0.300	0.319	4	4	8	1.422	1.434	0.002

Table 1 (Continued)

no.	molecule	bond	p_P	p_{cr}	n	m	l	d_{exp}/Å	d_{calc}/Å	σ/Å
223		q	0.400	0.377	4	3	6	1.422	1.420	0.002
224	benz[*a*]anthracene	a	0.143	0.130	4	3	6	1.483	1.471	0.011
225		b	0.428	0.404	3	2	5	1.401	1.404	0.012
226		c	0.571	0.601	2	1	3	1.400	1.356	0.011
227		d	0.428	0.452	2	1	2	1.392	1.383	0.014
228		e	0.571	0.601	2	1	3	1.393	1.356	0.014
229		f	0.428	0.404	3	2	4	1.418	1.403	0.011
230		g	0.428	0.404	4	2	5	1.442	1.414	0.012
231		h	0.143	0.156	3	2	4	1.396	1.453	0.013
232		i	0.857	0.899	2	1	4	1.322	1.302	0.011
233		j	0.143	0.130	3	2	4	1.429	1.455	0.012
234		k	0.286	0.268	4	2	5	1.434	1.439	0.012
235		l	0.571	0.571	3	2	5	1.384	1.375	0.011
236		m	0.428	0.428	3	2	5	1.431	1.401	0.013
237		n	0.286	0.268	4	2	4	1.397	1.439	0.013
238		o	0.286	0.306	3	2	4	1.436	1.426	0.012
239		p	0.714	0.750	2	1	3	1.323	1.329	0.014
240		q	0.286	0.306	2	1	2	1.444	1.409	0.015
241		r	0.714	0.750	2	1	3	1.360	1.329	0.013
242		s	0.286	0.306	3	2	4	1.428	1.426	0.013
243		t	0.428	0.404	3	2	5	1.422	1.404	0.011
244		u	0.571	0.541	3	2	6	1.364	1.378	0.012
245	2,3,8,9-dibenzo-	a	0.200	0.185	4	3	6	1.458	1.461	0.032
246	perylene	b	0.400	0.377	4	3	7	1.384	1.425	0.032
247		c	0.400	0.400	3	2	5	1.422	1.407	0.032
248		d	0.600	0.631	2	1	3	1.387	1.350	0.032
249		e	0.400	0.423	2	1	3	1.381	1.388	0.032
250		f	0.600	0.600	3	2	5	1.394	1.370	0.032
251		g	0.000	−0.007	4	3	7	1.478	1.497	0.032
252		h	0.400	0.377	4	3	5	1.454	1.425	0.032
253		i	0.200	0.185	4	3	6	1.479	1.461	0.032
254		j	0.800	0.761	3	2	7	1.406	1.337	0.032
255		k	0.200	0.215	3	2	5	1.409	1.442	0.032
256		l	0.400	0.377	4	2	3	1.379	1.418	0.032
257		m	0.400	0.423	3	2	3	1.413	1.404	0.032
258		n	0.600	0.631	2	1	2	1.384	1.350	0.032
259		o	0.400	0.423	2	1	2	1.399	1.388	0.032
260		p	0.600	0.631	2	1	3	1.412	1.350	0.032
261		q	0.400	0.423	3	2	3	1.403	1.404	0.032
262	1,14-benzobis-	a	0.533	0.562	2	1	2	1.40	1.363	0.02
263	anthrene	b	0.467	0.467	3	2	4	1.39	1.394	0.02
264		c	0.400	0.377	4	3	6	1.420	1.425	0.02
265		d	0.133	0.145	3	2	4	1.46	1.454	0.02
266		e	0.867	0.969	2	1	4	1.35	1.295	0.02
267		f	0.133	0.145	3	2	4	1.47	1.454	0.02
268		g	0.233	0.217	4	3	6	1.44	1.455	0.02
269		h	0.633	0.666	3	2	5	1.37	1.361	0.02
270		i	0.367	0.367	3	2	5	1.40	1.413	0.02
271		j	0.300	0.281	4	3	6	1.42	1.443	0.02
272		k	0.333	0.353	3	2	4	1.43	1.417	0.02
273		l	0.667	0.701	2	1	3	1.37	1.338	0.02
274		m	0.333	0.353	2	1	3	1.43	1.401	0.02
275		n	0.667	0.701	3	2	5	1.36	1.354	0.02
276		o	0.300	0.281	4	3	7	1.43	1.443	0.02
277		p	0.033	0.041	4	3	6	1.49	1.490	0.02
278		q	0.400	0.377	4	4	8	1.40	1.431	0.02
279		r	0.133	0.121	4	4	8	1.47	1.479	0.02
280		s	0.467	0.441	4	4	8	1.41	1.419	0.02
281		t	0.300	0.281	4	4	8	1.43	1.449	0.02
282		u	0.300	0.281	4	4	8	1.44	1.449	0.02
283	ovalene	a	0.00	0.185	4	4	8	1.435	1.467	0.006
284		b	0.400	0.377	4	4	8	1.415	1.431	0.004
285		c	0.300	0.281	4	3	6	1.424	1.443	0.004
286		d	0.500	0.500	3	2	6	1.400	1.388	0.004
287		e	0.200	0.215	3	2	4	1.441	1.442	0.004
288		f	0.800	0.839	2	1	4	1.356	1.313	0.004
289		g	0.200	0.215	3	2	4	1.429	1.442	0.004
290		h	0.400	0.423	4	3	6	1.450	1.421	0.004
291		i	0.400	0.423	3	2	4	1.413	1.404	0.004
292		j	0.600	0.631	2	1	4	1.365	1.351	0.006
293		k	0.300	0.281	4	4	8	1.413	1.449	0.006
294		l	0.300	0.281	4	4	8	1.411	1.449	0.004
295	tetrabenzo[*de*,*no*,	a	0.809	0.849	2	1	4	1.35	1.311	0.02
296	*st*,*c*′,*d*′]heptacene	b	0.191	0.206	3	2	4	1.45	1.443	0.02

Table 1 (Continued)

no.	molecule	bond	p_P	p_{cr}	n	m	l	d_{exp}/Å	d_{calc}/Å	σ/Å
297		c	0.182	0.168	4	3	4	1.42	1.464	0.02
298		d	0.427	0.451	3	2	3	1.39	1.399	0.02
299		e	0.573	0.603	2	1	3	1.39	1.356	0.02
300		f	0.427	0.451	2	1	4	1.39	1.383	0.02
301		g	0.573	0.573	3	2	5	1.38	1.375	0.02
302		h	0.382	0.360	4	3	5	1.44	1.428	0.02
303		i	0.236	0.220	4	4	4	1.42	1.460	0.02
304		j	0.045	0.036	4	3	6	1.48	1.489	0.02
305		k	0.227	0.211	4	2	6	1.44	1.450	0.02
306		l	0.727	0.691	3	2	6	1.37	1.350	0.02
307		m	0.273	0.291	3	2	5	1.42	1.428	0.02
308		n	0.227	0.211	4	2	4	1.42	1.450	0.02
309		o	0.500	0.527	3	2	4	1.38	1.385	0.02

[a] The bonds are numbered according to molecular graphs in Figures 1 and 2. The bonds 1−223 are from the training/validation set and 224−309 from the prediction set. The molecular descriptors are as follows: p_P − Pauling π-bond order, p_{cr} − Pauling π-bond order including corrections for the crystal packing effects calculated following the complicated scheme only for the prediction set, n − the number of neighboring carbon atoms around the bond, m − the number of benzenoid rings around the bond, l − the number of neighboring carbon atoms around those atoms counted for n, d_{exp} − the bond lengths from X-ray or neutron structure determinations (in Å), d_{calc} − (the model 8 was used) calculated (for the training/validation set) or predicted (for the prediction set) d's (in Å), σ − estimated standard deviations for d_{exp} (in Å).

Table 2. Bond Length−Bond Descriptor Correlations[a]

p_P:	**−0.895**	p_P, p_P2:	0.898	log(1 + p_P):	−0.898	f_P:	−0.896
p_{cr}:	**−0.929**	p_{cr}, $p_{cr}2$:	0.931	log(1 + p_{cr}):	−0.931	f_{cr}:	−0.927
n:	**0.735**	m:	**0.689**	l:	**0.502**	k:	0.250
n, n^2:	0.741	m, m^2:	**0.748**	l, l^2:	0.506	l, ..., l^6:	**0.630**

[a] The significant correlations used in regression models are bold. Pauling curve is defined as $f_P = 1.84\ p_P/(0.84\ p_P + 1)$ and $f_{cr} = 1.84\ p_{cr}/(0.84\ p_{cr} + 1)$.

Table 3. PCA Results for the Training/Validation Set

PC	% variance	% cumul. variance	p_P	p_{cr}	n	m	l
PC1	74.08	74.08	−0.414	−0.430	0.482	0.482	0.424
PC2	22.38	96.46	0.567	0.527	0.206	0.292	0.522
PC3	2.55	99.01	0.128	0.057	0.844	−0.432	−0.286
PC4	0.85	99.87	0.147	0.088	0.101	0.704	−0.682
PC5	0.13	100.00	−0.685	0.725	0.058	0.021	−0.024

tion of d, the models 7−10, reach up-to-date experimental precision ≈0.005 Å (comparable to $\langle\Delta\rangle$ values), and are relatively far from the limit $\langle\Delta/\sigma\rangle = 2.58$.

The regression vector for model 8 shows almost equal contribution of the bond orders and less contribution of the topological indices in linear decreasing order $n - m - l$:

$$p_P: -0.357,\ p_{cr}: -0.385,\ n:\ 0.157,\ m:\ 0.114,\ l:\ 0.006$$

The regression plot (Figure 6) shows interesting distribution of shortened, unchanged, and lengthened bonds. Shortened bonds (blue dots) are dominant in the range 1.33−1.40 Å. Lengthened bonds (red dots) in this region are regularly below the regression line (green). The region with the highest mixing degree is 1.40−1.45 Å. The last range 1.45−1.48 Å is dominant by red dots. Unchanged bonds (magenta dots) are practically not present in this region. These observations lead to the conclusion that most aromatic bonds which are longer or even formally single in a vacuum undergo lengthening in the crystal field. On the other side, short aromatic bonds and almost double bonds get shorter in the crystal in most cases. By other words, in general, shorter bonds get stronger (closer to double bond) and longer bonds become weaker (more single in character) due to attractive intermolecular interactions, charge transfer and π-system adjustment in crystal.

The approximate prediction of d based on the n, m, l numbers (models 3−6) is less accurate than the models including the bond orders, but model 5 (containing all the information from the original variables) can be recommended as the fast, easy and approximate prediction of d's in PB−PAHs. The regression models exhibit that nonlinear forms of m and l, i.e. m_{cr} and l_{cr}, are needed whenever the bond orders are not included.

The approximate predictions of p_P and p_{cr} based on the n, m, l (models 11−18) behave as the analogous models for prediction of d. The proposed models are those containing the maximum information (all PCs) models, models 14 and 18. In such a case all the information is required, as observed when the corresponding MLR and PLS are compared.

Model 8 was used to predict d's of the prediction set (Table 1). The experimental values were compared with predicted values: $R = 0.835$, $SEP = 0.016$ Å, $\langle\Delta\rangle = 0.021$ Å, $\langle\Delta/\sigma\rangle = 1.634$. As the experiments are inaccurate or old, even this comparison shows that model 8 is good and applicable for our purposes.

f. Structural Aromaticity Indices. Table 5 contains data for Julg's structural aromaticity index and average bond length. The problem that arises here, and basically is present in the whole work, is which bond length to use: corrected to thermal motion in crystal, or uncorrected? To average them to get unique bonds or not? It seems that there is not significant change if choosing any of these options. For example, it is always $A = 1.000$ for benzene. If the d's are not averaged and not corrected to thermal motion $A = 0.999(17)$ (structure: BENZENE). Another example: d's of hexabenzocoronene (structure: HBZCOR01). If the bond lengths are not averaged and not corrected to thermal motion $A = 0.907(100)$, and if they are just corrected to thermal motion $A = 0.907(100)$. If both averaging and the thermal motion corrections are applied then $A = 0.909(101)$. There is no significant difference. Thermal corrections are usually up to 0.003 Å, and of the same order are the differences between bond lengths that are equal in gas phase but in crystal are different due to lower molecular symmetry. Lewis and Peters[21] pointed out that 0.01 Å is the accuracy of

Table 4. Regression Models[a]

no.	Y	bond descriptors	method	R	Q	SEV	$\langle\Delta\rangle$	$\langle\Delta/\sigma\rangle$
1[b]	d/Å	p_P	LR	0.895	0.893	0.014	0.010	3.345
2[b]	d/Å	p_{cr}	LR	0.929	0.929	0.011	0.008	2.344
3[b]	d/Å	n, m, l	MLR	0.836	0.830	0.017	0.014	4.492
4	d/Å	n, m, l	PLS	0.820	0.814	0.018	0.014	4.603
5[b]	d/Å	n, m, l, m_{crd}, l_{crd}	MLR	0.848	0.839	0.016	0.013	4.142
6[c]	d/Å	n, m, l, m_{crd}, l_{crd}	PLS	0.838	0.830	0.017	0.013	4.195
7	d/Å	p_P, p_{cr}, n, m, l	MLR	0.959	0.957	0.009	0.006	1.823
8	d/Å	p_P, p_{cr}, n, m, l	PLS	0.940	0.938	0.011	0.007	2.167
9	d/Å	p_P, p_{cr}, n, m, l, m_{crd}, l_{crd}	MLR	0.960	0.957	0.009	0.006	1.749
10	d/Å	p_P, p_{cr}, n, m, l, m_{crd}, l_{crd}	PLS	0.943	0.941	0.010	0.007	2.058
11[b]	p_P	n, m, l	MLR	0.795	0.787	0.115	0.091	
12	p_P	n, m, l	PLS	0.765	0.757	0.122	0.098	
13	p_P	n, m, l, m_{crpp}, l_{crpp}	MLR	0.800	0.789	0.114	0.089	
14[c]	p_P	n, m, l, m_{crpp}, l_{crpp}	PLS	0.779	0.767	0.119	0.094	
15[b]	p_{cr}	n, m, l	MLR	0.815	0.809	0.111	0.092	
16	p_{cr}	n, m, l	PLS	0.789	0.782	0.117	0.099	
17	p_{cr}	n, m, l, m_{crpcr}, l_{crpcr}	MLR	0.822	0.812	0.110	0.089	
18[c]	p_{cr}	n, m, l, m_{crpcr}, l_{crpcr}	PLS	0.802	0.792	0.115	0.095	

[a] The Linear Regression (LR), Multiple Linear Regression (MLR), and Partial Least Squares (PLS) models for prediction of d, p_P, and p_{cr}. The corrected variables have the form $m_{cr} = m + am^2$ and $l_{cr} = l + al^2 + bl^3 + cl^4 + dl^5 + el$,[6] where the coefficients $a-e$ are found in the polynomial fitting to d, p_P, and p_{cr} (indexes of the corrections are marked as crd, crpp, and crpcr, respectively). R and Q are the prediction and validation correlation coefficients, *SEV* is the standard error of validation, $\langle\Delta\rangle$ is the average absolute deviation of predicted d's from experimental, $\langle\Delta/\sigma\rangle$ is the average Δ/σ ratio where σ is the experimental estimated standard deviation on d's. *SEV* and $\langle\Delta\rangle$ are in Å when referred to bond lengths d.[b] All the regression coefficients are greater than their statistical errors more than 2.58 times. The model 5 hardly satisfy this condition. [c] The models with three PCs used. Other PLS models are performed with two PCs.

Table 5. Experimental and Predicted Structural Aromaticity Indices[a]

molecule	A_{exp}	A_{M1}	A_{M2}	A_{M8}	$\langle d_{exp}\rangle$	$\langle d_{M1}\rangle$	$\langle d_{M2}\rangle$	$\langle d_{M8}\rangle$
benzene	1.000	1.000	1.000	1.000	1.390	1.394	1.380	1.384
naphthal.	0.932(41)	0.928(42)	0.920(45)	0.920(45)	1.401	1.401	1.402	1.397
anthracene	0.889(119)	0.878(125)	0.863(132)	0.884(119)	1.400	1.403	1.401	1.400
phenanth.	0.878(215)	0.928(154)	0.908(173)	0.907(171)	1.393	1.404	1.402	1.400
tetracene	0.870(77)	0.849(83)	0.843(83)	0.870(77)	1.403	1.405	1.405	1.403
triphenyl.	0.906(283)	0.946(214)	0.940(225)	0.925(258)	1.407	1.405	1.406	1.404
chrysene	0.848(84)	0.922(60)	0.906(66)	0.906(65)	1.400	1.405	1.408	1.408
pyrene	0.916(120)	0.899(129)	0.899(129)	0.896(131)	1.401	1.406	1.405	1.405
perylene	0.877(96)	0.890(87)	0.882(92)	0.886(91)	1.404	1.407	1.406	1.406
1,2,5,6-da.	0.872(85)	0.906(72)	0.889(79)	0.900(75)	1.404	1.405	1.405	1.404
picene	0.900(324)	0.921(288)	0.918(294)	0.912(271)	1.404	1.406	1.405	1.404
3,4-benzp.	0.877(473)	0.890(468)	0.884(421)	0.888(471)	1.404	1.407	1.406	1.406
pentacene	0.880(224)	0.826(360)	0.837(261)	0.862(240)	1.410	1.405	1.407	1.405
dbanthrac.	0.891(39)	0.915(34)	0.904(37)	0.904(37)	1.408	1.405	1.405	1.403
dbtetracene	0.881(213)	0.944(147)	0.939(153)	0.922(176)	1.418	1.407	1.408	1.408
bzperylene	0.875(299)	0.921(251)	0.911(266)	0.903(274)	1.407	1.408	1.406	1.408
coronene	0.955(134)	0.933(163)	0.945(147)	0.923(174)	1.409	1.409	1.410	1.411
dbphpenta.	0.866(140)	0.936(97)	0.928(102)	0.915(117)	1.409	1.410	1.409	1.411
quaterryl.	0.889(216)	0.877(227)	0.877(230)	0.880(224)	1.411	1.409	1.410	1.411
hbcoronene	0.910(101)	0.948(77)	0.940(82)	0.924(93)	1.411	1.410	1.413	1.415
kekulene	0.877(124)	0.881(122)	0.880(123)	0.893(116)	1.409	1.409	1.409	1.409
bdcoronene	0.925(98)	0.921(101)	0.923(99)	0.913(106)	1.413	1.411	1.411	1.415

[a] Julg's aromaticity index based on experimental (A_{exp}) and calculated bond lengths from models 1 (A_{M1}), 2 (A_{M2}), and 8 (A_{M8}) from Table 4. Errors are in brackets, given at last 2−3 digits. Average bond lengths (in Å) from experiment ($\langle d_{exp}\rangle$) and from models 1, 2, and 8 ($\langle d_{M1}\rangle$, $\langle d_{M2}\rangle$, $\langle d_{M8}\rangle$, respectively). Their errors are at most 0.001 Å.

measuring bond lengths without vibrations. For benzene, if vibrations are included, accuracy is 0.1 Å. This seriously puts in question any bond length prediction. But let us give a simple example. Benzene solvate at 10 K (neutron diffraction, structure PPRHZ01) has average $\langle d\rangle = 1.400(6)$ Å and $A = 0.986(1)$. Strictly speaking, it could be said that $\langle d\rangle = 1.400(2)$ Å (σ averaged as in accompanied expression for A) and the packing effect on bond lengths, calculated by a method after Bürgi,[32] is 0.011 Å. Then, the composite accuracy of the bond length measurement at zero level is rounded to 0.01 Å, which is the order of precision of some models in this work (4 from 8 models have $\Delta \approx 0.01$ Å, Table 4).

There is no significant difference between experimental and any predicted A due to large errors originating from $\sigma > 0.001$ Å for bond lengths. The $\langle d\rangle$ values have $\sigma \leq 0.001$ Å, and so they differ significantly in some cases. Correlation of experimental with predicted aromaticity indices can be some measure of the model quality. This way model 8 is better than model 2 (which is better than model 1) when A is considered (correlations with experimental values, $r =$ 0.761; 0.707; 0.590, respectively). Models 8 and 2 are better than 1, but 2 is better than 8 when $\langle d\rangle$ is used as a model quality parameter ($r = 0.823$; 0.850; 0.768, respectively). Anyway, the multivariate model or the univariate one with two-variable function as p_{cr} gives better results than pure

univariate model (with p_P).

The Julg's index A shows that benzene is the most aromatic hydrocarbon, while the others are almost at the same level. The least aromatic is chrysene according to A_{exp}, but predictions suggested pentacene, which is to be expected due to large differences between alternating terminal bonds.

If the regression equation $\langle d \rangle = a + b/m_r$ is established, where m_r is the number of hexagonal rings in a molecule, models 1, 2, 8 give much better regression models than the experiment ($r = 0.954$; 0.938; 0.937; 0.766, respectively). The a coefficient is 1.413 Å for model 8, 1.410 Å for model 1, and 1.411 Å for other two models. These values of a are on the halfway from benzene (1.400 Å) to graphite (1.422 Å). It means that by increasing the size of PB−PAH molecule the average $\langle d \rangle$ increases, so that for infinite size it would reach a. Large PB−PAH molecule, much larger than deposited in CSD, would consist of graphitic like bonds in the molecular interior, alternated bonds at molecular exterior (bonds with hydrogens) and in its closest neighborhood, and bonds with medium length in the regions between the interior and exterior.

5. CONCLUSIONS

At the end of this study the answers on questions from the introductory part can be given. Besides, there are some additional conclusions coming out from the results. The aromatic carbon−carbon bond in PB−PAHs from crystal structures:

I − is at least two-dimensional phenomenon. PC1 is connected to bond length and PC2 to the shape of the bond neighborhood, which is equal to the position of the bond in the molecule.

II − depends both on bond orders (p_P, p_{cr}) and topological indices (n, m, l, m_{cr}, l_{cr}) describing the bond neighborhood.

III − can be classified in 12 classes with distinguished n and $n + m + l$ numbers.

IV − is better predicted if crystal effects are introduces (p_{cr} or its functions).

V − is better predicted with multivariate models based on PCs than with univariate. The model recommended here is the PLS model with p_P, p_{cr}, n, m, l as bond descriptors.

VI − contains some information which is difficult to rationalize,[36] as crystal structures cannot be predicted up-to-date, which limits the full quantification of crystal effect corrections. Besides, the bond length prediction in this work, compared to experimental errors, is accurate enough for various studies.

VII − can be, as well as p_P and p_{cr}, predicted approximately using regression models with topological indices only (n, m, l, m_{cr}, l_{cr}), avoiding this way the use of complicated algorithms and calculations (the recommended models are MLR or PLS models with all the PCs).

VIII − is predicted for the prediction set satisfactorily well.

IX − is a good starting point for the study of other aromatic bonds (C−N, C−O, etc.) in different molecular classes.

X − shows that the bond orders and the topological indices are not mutually orthogonal but exhibit moderate correlation ($r = 0.33-0.64$).

XI − when predicted with a more accurate model usually gives a better prediction of structural aromaticity indices.

XII − indicates that there are two standards for aromatic hydrocarbons around which the bond lengths tend to cluster: benzene and graphite, with maximum (two) and minimum (zero) number of hydrogens bound to aromatic bond, respectively.

ACKNOWLEDGMENT

The authors acknowledge FAPESP for the financial support.

APPENDIX: THE LIST OF CSD REFCODES WITH REFERENCES

ANTCEN10: Brock, P. C.; Dunitz, J. D. Temperature-dependence of Thermal Motion in Crystalline Anthracene. *Acta Crystallogr*. **1990**, *B46*, 795−806.

BEANTH: Trotter, J. The Crystal Structure of 1−14-Benzbisanthrene. *Acta Crystallogr*. **1958**, *11*, 423−428.

BENZEN: Bacon, G. E.; Curry, N. A.; Wilson, S. A. A Crystallographic Study of Solid Benzene by Neutron Diffraction. *Proc. R. Soc. London* **1964**, *A279*, 98−110.

BNPERY02: Munakata, M.; Wu, L. P.; Ning, G. L.; Kuroda-Sowa, T.; Maekawa, M.; Suenaga, Y.; Maeno, N. Construction of metal sandwich systems derived from assembly of silver(I) complexes with polycyclic aromatic compounds. *J. Am. Chem. Soc*. **1999**, *121*, 4968−4976.

BNPYRE10: Iball, J.; Scrimgeour, S. N.; Young, D. W. 3,4-Benzopyrene (A New Refinement). *Acta Crystallogr*. **1976**, *B32*, 328−330.

BZAPRM10: Foster, R.; Iball, J.; Scrimgeour, S. N.; Williams, B. C. Crystal-Structure and Nuclear Magnetic-Resonance Spectra of a 1−1 Complex of Benz[*a*]anthracene and Pyromellitic Dianhydride (Benzene-1,2,4,5-tetracarboxylic Dianhydride). *J. Chem. Soc., Perkin Trans. 2* **1976**, 682−685.

CRYSEN01: Krygowski, T. M.; Ciesielski, A.; Swirska, B.; Leszczynski, P. 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. *Pol. J. Chem*. **1994**, *68*, 2097−2107.

DBPERY: Lipscomb, W. N.; Robertson, J. M.; Rossmann, M. G. Crystal-Structure Studies of Polynuclear Hydrocarbons. 1. 2−3−8−9-Dibenzoperylene. *J. Chem. Soc*. **1959**, 2601−2607.

DBNTHR02: Iball, J.; Morgan, C. H.; Zacharias, D. E. Refinement of Crystal-structure of Orthorhombic Dibenz-[*a*,*h*]-anthracene. *J. Chem. Soc., Perkin Trans. 2* **1975**, 1271−1272.

DUPCIA10: Mitani, S. Crystal structure of coronene iodide. *Annu. Rep. Res. R. I. Kyoto U*. **1986**, *20*, 125−130.

FILHOX: Bulgarovskaya, I.; Zavodnik, V. E.; Vozzhennikov, V. M. Structure of the 1:1 π-Molecular Complex of Tetracene with 1,2:4,5-Pyromellitic Dianhydride. *Acta Crystallogr*. **1987**, *C43*, 764−766.

HBZCOR01 Goddard, R.; Haenel, M. W.; Herndon, W. C.; Krüger, C.; Zander, M. 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′*]dicoronene. *J. Am. Chem. Soc.* **1995**, *117*, 30−41.

KEKULN10: Staab, H. A.; Diederich, F.; Krieger, C.; Schweitzer, D. Molecular Structure and Spectroscopic Properties of Kekulene. *Chem. Ber*. **1983**, *116*, 3504−3512.

NAPHTA10: Brock, C. P.; Dunitz, J. D. Temperature Dependence of Thermal Motion in Crystalline Naphthalene. *Acta Crystallogr*. **1982**, *B38*, 2218−2228.

OVALEN01: Hazell, R. G.; Pawley, G. S. The crystal and molecular structure of ovalene: some threee-dimensional refinements. *Z. Kristallogr*. **1975**, *137*, 159−172.

PENCEN01: Holmes, D.; Kumaraswamy, S.; Matzger, A. J.; Vollhardt. K. P. C. On the nature of nonplanarity in the [*N*]phenylenes. *Chem.-Eur. J*. **1999**, *5*, 3399−3412.

PERLEN03: Krygowski, T. M.; Ciesielski, A.; Swirska, B.; Leszczynski, P. 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. *Pol. J. Chem*. **1994**, *68*, 2097−2107.

PHENAN08: Petricek, V.; Cisarova, I.; Hummel, L.; Kroupa, J.; Brezina, B. Orientational Disorder in Phenanthrene. Structure Determination at 248, 295, 339 and 344 K. *Acta Crystallogr*. **1990**, *B46*, 830−832.

PINZOB: Boje, B.; Magull, J. On the Reaction of Dilithiumbiphenyl with $SmBr_3$ − The Crystal Structure of $[(C_{24}H_{16})SmBr(THF)_2]_2$...$[C_{24}H_{14}]$. *Z. Anorg. Allg. Chem*. **1994**, *620*, 703−706.

PPCRHZ01 Boucherle, J. X.; Gillon, B.; Maruani, J.; Schweizer J. Crystal Structure Determination by Neutron Diffraction of 2,2-Diphenyl-1-picrylhydrazyl (DPPH) Benzene Solvate (1/1). *Acta Crystallogr*. **1987**, *43*, 1769−1773.

PYRENE01: Allmann, R. Significance of Weak and Unobserved Reflexes in Structural Refining-Pyrene, $C_{16}H_{10}$. *Z. Kristallogr*. **1970**, *132*, 129−151.

PYRENE02: Hazell, A. C.; Larsen, F. K.; Lehmann, M. S. A Neutron Diffraction Study of the Crystal Structure of Pyrene, $C_{16}H_{10}$. *Acta Crystallogr*. **1972**, *B28*, 2977−2984.

QUATER10: Kerr, K. A.; Ashmore, J. P.; Speakman, J. C. Crystal and Molecular Structure of Quaterrylene − Redetermination. *Proc. R. Soc. London* **1975**, *A344*, 199−215.

RULLUF: Carrell, H. L.; Glusker, J. P. The Molecular Complex of Dibenz[a,c]anthracene and Trinitrobenzene. *Struct. Chem*. **1997**, *8*, 141−147.

TBZHCE: Ferguson, G.; Parvez, M. Tetrabenzo-[*de*,*no*,*st*,c_1d_1]heptacene. *Acta Crystallogr*. **1979**, *B35*, 2419−2421.

TRIPHE11: Ferraris, G.; Jones, D. W.; Yerkess, Y. A Neutron-diffraction study of the crystal and molecular structure of triphenylene. *Z. Kristallogr*. **1973**, *138*, 113−128.

WOQPAT: Kubel, C.; Eckhardt, K.; Enkelmann, V.; Wegner, G.; Mullen, K. Synthesis and crystal packing of large polycyclic aromatic hydrocarbons: hexabenzo[*bc*, *ef*, *hi*, *kl*, *no*, *qr*]coronene and dibenzo[*fg*, *ij*]phenanthro[9, 10, 1, 2, 3-*pqrst*]pentaphene. *J. Mater. Chem*. **2000**, *10*, 879−886. This structure was not found in Cambridge Structural Database October 2000 Release which was the database version used in this work but was found recently in the next CSD April 2001 Release.

YOFCUR: Goddard, R.; Haenel, M. W.; Herndon, W. C.; Krüger, C.; Zander, M. 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′*]dicoronene. *J. Am. Chem. Soc*. **1995**, *117*, 30−41.

ZZZOYC01: De, A.; Ghosh, R.; Roychowdhury, S.; Roychowdhury, P. Structural Analysis of Picene, $C_{22}H_{14}$. *Acta Crystallogr*. **1985**, *C41*, 907−909.

REFERENCES AND NOTES

(1) Allen, F. H.; Kennard, O.; Watson, D. G.; Brammer, L.; Orpen, A. G. 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.

(2) Kiralj, R. Structural Studies of 4,9-Diazapyrene Derivatives; Ph.D. University of Zagreb, Zagreb, Croatia, 1999.

(3) Burley, S. K.; Petsko, G. A. Aromatic−Aromatic Interaction: A Mechanism of Protein Structure Stabilization. *Science* **1985**, *229*, 23−28.

(4) Trucano, P.; Chen, R. Structure of graphite by neutron diffraction. *Nature* **1975**, *258*, 136−137.

(5) Herndon, W. C.; Párkányi, C. π Bond Orders and Bond Lengths. *J. Chem. Educ*. **1976**, *53*, 689−691.

(6) Narita, S.; Morikawa, T.; Shibuya, T. Linear relationship between the bond lengths and the Pauling bond orders in fullerene molecules. *J. Mol. Struct. (THEOCHEM)* **2000**, *532*, 37−40.

(7) Pauling, L. Bond Numbers and Bond Lengths in Tetrabenzo[*de*, *no*, *st*, c_1d_1]heptacene and Other Condensed Aromatic Hydrocarbons: A Valence-Bond Treatment. *Acta Crystallogr*. **1980**, *B36*, 1898−1901.

(8) Morikawa, T.; Narita, S.; Shibya, T.-I. Graph-theoretical rules for predicting bond lengths in hexagonal-shaped benzenoid hydrocarbons. *J. Mol. Struct. (THEOCHEM)* **1999**, *466*, 137−143.

(9) Goddard, R.; Haenel, M. W.; Herndon, W. C.; Krüger, C.; Zander, M. 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′*]dicoronene. *J. Am. Chem. Soc*. **1995**, *117*, 30−41.

(10) Galvez, J. On a topological interpretation of electronic and vibrational molecular energies. *J. Mol. Struct. (THEOCHEM)* **1998**, *429*, 255−264.

(11) Herndon, W. C. Resonance Theory. VI. Bond Orders. *J. Am. Chem. Soc*. **1974**, *96*, 7605−7614.

(12) Dewar, M. J. S.; Gleischer, G. J. Ground States of Conjugated Molecules. VII. Compounds Containing Nitrogen and Oxygen. *J. Chem. Phys*. **1966**, *44*, 759−773.

(13) Dewar, M. J. S.; Llano, C. Ground States of Conjugated Molecules. XI. Improved Treatment of Hydrocarbons. *J. Am. Chem. Soc*. **1969**, *91*, 789−795.

(14) Kiralj, R.; Kojić-Prodić, B.; Žinić, M.; Alihodžić, S.; Trinajstiæ, N. 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-$\kappa^5O,O',O'',O''',O''''$](picrate-$\kappa^2O,O'$)potassium, and [*N*,*N*′-Bis(6-phenanthridinyl-κN-methyl)-7,16-diaza-18-crown-6-κ^4O,O',O'',O'''] sodium Iodide Dichloromethane Solvate. *Acta Crystallogr*. **1996**, *B52*, 823−837.

(15) Kiralj, R.; Kojić-Prodić, B.; Nikolić, S.; Trinajstić, N. Bond lengths and bond orders in benzenoid hydrocarbons and related systems: a comparison of valence bond and molecular orbital treatments. *J. Mol. Struct. (THEOCHEM)* **1998**, *427*, 25−37, and the references therein.

(16) Kiralj, R.; Kojić-Prodić, B.; Piantanida, I.; Žinić, M. Crystal and molecular structures of diazapyrenes and a study of π...π interactions. *Acta Crystallogr*. **1999**, *B55*, 55−69.

(17) Lendvay, G. On the correlation of bond order and bond length. *J. Mol. Struct. (THEOCHEM)* **2000**, *501−502*, 389−393.

(18) Schleyer, P. R.; Freeman, P. K.; Jiao, H.; Goldfuss, B. Aromaticity and Antiaromaticity in Five-Membered C_4H_4X Ring Systems: "Classical" and "Magnetic" Concepts May Not Be "Orthogonal". *Angew. Chem., Int. Ed. Engl*. **1995**, *34*, 337−340.

(19) Schleyer, P. R.; Jiao, H. What is Aromaticity? *Pure Appl. Chem*. **1996**, *68*, 209−218.

(20) Glukhotsev, M. Aromaticity Today: Energetic and Structural Criteria. *J. Chem. Educ*. **1997**, *74*, 132−136.

(21) Lewis, D.; Peters, D. *Facts and Theories of Aromaticity*, 1st ed; The MacMillan Press Ltd.: London, 1975; pp 10−14.

(22) Krygowski, T. M.; Ciesielski, A.; Bird, C. W.; Kotschy, A. Aromatic Character of the Benzene Ring Present in Various Topological Environments in Benzenoid Hydrocarbons. Nonequivalence of Indices of Aromaticity. *J. Chem. Inf. Comput. Sci*. **1995**, *38*, 203-210.

(23) Allen, F. K.; Kennard, O. 3D Search and Research Using the Cambridge Structural Database. *Chem. Design Autom. News* **1993**, *8*, 131−137.
(24) Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. *Chemometrics: A Practical Guide*; Wiley: New York, 1998
(25) Beebe, K. R.; Kowalski, B. P. An Introduction to Multivariate Calibration and Analysis. *Anal. Chem*. **1987**, *59*, 1007A−1017A.
(26) Cambridge Structural Database, October Release 2000. The Chemistry Visualization Program (NCSA ChemViz) at the National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign. http://chemviz.ncsa.uiuc.edu
(27) Trinajstić, N. *Chemical Graph Theory*, 2nd ed.; CRD Press: Boca Raton, FL, 1992.
(28) Randić, M. Graph Theoretical Derivation of Pauling Bond Orders. *Croat. Chem. Acta* **1987**, *47*, 71−78.
(29) Stoicheff, B. P. Variation of Carbon−Carbon Bond Lengths With Environment as Determined by Spectroscopic Studies of Simple Polyatomic Molecules. *Tetrahedron* **1962**, *17*, 135−145.
(30) Vilkov, L. V.; Mastryukov, V. S.; Sadova, N. I. *Determination of the Geometrical Structure of Free Molecules*; Mir Publishers: Moscow, 1983; pp 88−89.
(31) Kiralj, R. Structural Investigations of Macrocyclic Receptors with Phenanthridine Subunits; Master Thesis, University of Zagreb, Zagreb, Croatia, 1994.
(32) Bürgi, H. B. Structure correlation and chemistry. *Acta Crystallogr*. **1998**, *A54*, 873−885.
(33) Glusker, J. P.; Lewis, M.; Rossi, M. *Crystal Structure Analysis for Chemists and Biologists*; VCH Publishers: New York, 1994; pp 429−431.
(34) Matlab 5.4; The MathWorks, Inc.: Natick, MA, 2001.
(35) Pirouette 3.01; Infometrix, Inc.: Seattle, WA, 2001.
(36) Gavezzotti, A. Are Crystal Structures Predictable? *Acc. Chem. Res*. **1994**, *27*, 309−314.

CI010063G