*Journal of Computer-Aided Molecular Design*, **14:** 679–687, 2000.
*KLUWER/ESCOM*



# Indices of differences of path lengths: Novel topological descriptors derived from electronic interferences in graphs

Jorge Gálvez*, Ramón Garcia-Domenech & Carolina de Gregorio-Alapont
*Molecular Connectivity and Drug Design Research Unit, Department of Physical Chemistry, Faculty of Pharmacy, University of Valencia, Avda. V.A. Estellés, s/n, E-46100 Burjassot, Valencia, Spain*

Received 3 June 1999; Accepted 9 March 2000

*Key words:* differences of path lengths, topological descriptors, wave interferences

**Summary**

Novel topological descriptors, namely indices of differences of path lengths (DPs), are deduced from the physical model of wave interferences. Two electrons, moving through a circuit graph within a diffraction experiment, interfere in a given vertex of the graph. It is demonstrated that the overall sum of the inverse of the squares of the differences of topological distances between all pairs of vertices of the graph is a measure of the mean global kinetic energy of the electrons which are able to produce a constructive interference. New topological indices, namely *indices of differences of path lengths* are thus introduced as derived from such a diffraction pattern. These indices, according to the above expressed, should be a measure of the electron mobility within the molecule. As a consequence, a good prediction is to be expected for properties related to such mobility, such as resonance energy in aromatic hydrocarbons. Our results confirm that in fact, the resonance energies are well predicted by this means. Moreover, the new indices demonstrate to be very useful in the evaluation of biological properties such as antibacterial activities of a wide set of heterogeneous compounds.

## Introduction

Mathematical chemistry is becoming one of the most interesting areas of research in theoretical chemistry. However, most of its concepts are directly derived from mathematical formalisms such as Graph Theory or topological approaches. At present, the most frequent criticism and an essential disadvantage is the absence of any physical–chemical interpretation of the majority of this type of descriptors. Moreover, it is likely that there is no need for any physical or physicochemical interpretation of a topological (or graph theoretical) descriptor, since they are mathematical abstractions, but, on the contrary, there would be a topological interpretation of any physical or physicochemical magnitude. This disjunctive is a very interesting 'philosophical' question, closely related to the fuzzy notion of chemical structure. However, the interest of deriving topological descriptors from physicochemical magnitudes is obvious, since (see text) the physical profile able to be evaluated should be known.

There are some precedents in the deduction of topological indices (TIs) from physical models. The resistance distance introduced by Klein and Randic in 1993 [1] is a very good example of this kind of approach. They developed a mathematical model for graphs in which a resistance has been introduced in some edges of the graph. This kind of graph counts have proved their utility in the prediction of different molecular properties.

Moreover, in spite of the high number of TIs described in the literature, only a few of them, such as the complexity indices developed by Basak [2], the differences between valence and non-valence Kier and Hall connectivity indices [3], the E-state index [4,5] and the topological charge indices [6] are able to efficiently predict electronic properties. The importance of the topological (or graph theoretical) description

*To whom correspondence should be addressed. E-mail: Jorge.Galvez@uv.es

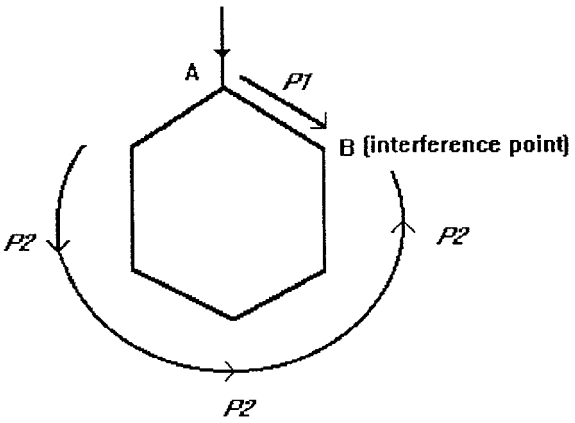


*Figure 1.* Interference between two electrons $e_1$ and $e_2$, moving through two different paths, $p_1$ and $p_2$, respectively. Here the difference of number of edges, or topological distances, between paths 1 and 2 is 4. Observe the point of interference (B).

of such properties in the design of new compounds showing predetermined properties, and especially new drugs, has been clearly demonstrated by the excellent efficiency shown by the topological charge indices [7–9].

Now, we introduce new topological descriptors, which seem to join two important advantages: First, they are deduced from a well-known physical model and, second, they encode significant electronic information. Thus, the new descriptors, namely 'indices of difference of lengths between paths' (DPs), are derived from the model of interference between electronic waves and they are able to predict electronic properties as, for example, resonance energies of conjugated hydrocarbons, but also other more complex properties such as microbiological ones.

## Theoretical formalism

The graph theoretical representation of benzene is an Euclidean 1-torus. Let us suppose that two different electrons, $e_1$ and $e_2$, are moving through two different paths, $p_1$ and $p_2$, as illustrated in Figure 1. In this figure we suppose that the molecule, i.e. benzene, is being shot with an electron beam just doing a diffraction experiment. This kind of experiment is well known since Davisson and Germer observed diffraction effects when an electron beam was reflected from a crystal of Ni. It is also well known that the wavelength of an electron moving at $10^8$ cm/s is about 7 Å, which is of the order of magnitude of molecular dimensions. Moreover, this kind of diffraction experiment has been reproduced even in the gas phase and from the Wierl equation the diffraction pattern of the molecule may be deduced and expressed as a function of the distances between all the possible pairs of atoms in the molecule [10]. In fact, electron diffraction by gases is a useful method in elucidating molecular structure given that, although the gas molecules have random orientations, the orientation of atoms with respect to one another is fixed, so that it is possible to determine structural parameters such as bond distances and angles.

Keeping in mind this model, we will consider as a first approach only the shortest path ($p_1$) and the second shortest path ($p_2$) between two electrons interfering in a diffraction experiment involving a simple aromatic molecule such as benzene.

Both electrons come from point A up to point B, where they interfere. The wave equations for the electrons $e_1$ and $e_2$ are $\Psi_1 = A \cdot \sin(\omega \cdot t_1 + \varphi)$ and $\Psi_2 = A \cdot \sin(\omega \cdot t_2 + \varphi)$, respectively, where $\Psi_1$ and $\Psi_2$ are the corresponding wavefunctions for electrons 1 and 2, respectively, $t_1$ and $t_2$ are the times necessary for electrons 1 and 2 to go from point A to B, respectively and $\varphi$ is the angle of initial phase.

Since both electrons will interfere in the vertex B, the resulting wave equation is:

$$\Psi_R = \Psi_1 + \Psi_2 = A_R \cdot \sin\left[\frac{\omega(t_1 + t_2)}{2} + \varphi\right] \quad (1)$$

where $\Psi_R$ is the resulting wavefunction, $\omega$ is the pulsation and $A_R$ is the resulting amplitude which is given by:

$$A_R = 2A \cdot \cos\left[\frac{\omega(t_1 - t_2)}{2}\right] \quad (2)$$

If we consider only the case of constructive interference, i.e. when $A_R = \pm 2A$, the following equality, $\cos\left[\frac{\omega(t_1-t_2)}{2}\right] = \pm 1$, implies that $\left[\frac{\omega(t_1-t_2)}{2}\right] = k\pi$, where $k$ is the interference order. The last expression can be written as: $\mathrm{v}(t_1 - t_2) = k \cdot \lambda$, where v is the electron velocity along the ring and $\lambda$ is the electron-associated wavelength. This expression can be transformed in the following by introducing the lengths $s_1$ and $s_2$ which electrons $e_1$ and $e_2$ have to follow up to the point B: $s_1 - s_2 = k \cdot \lambda$, and considering the distance $d$ as a mean value for the length of all the edges of the graph representing the circuit in which electrons interfere, we have $s_1 - s_2 = d(p_1 - P_2)$.

Then it follows that:

$$p_1 - p_2 = \frac{\lambda}{d} \quad \text{for} \quad k = 1 \quad (3)$$

*Table 1.* Comparison between 'experimental' resonance energy values and those obtained from Equation 10. All values are expressed in kcal.mol$^{-1}$. Values of the DP indices used in Equation 10 are also included

| Compound | $DP_1$ | $DP_2$ | $DP_3$ | $RE^a_{exp}$ | $RE^b_{calc}$ | Res. | $RE^c_{calc(CV)}$ | $Res_{(CV)}$ | $RE^d_{calc}$ | $RE^e_{calc}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Benzene | 0.375 | 1.500 | 0.000 | 36 | 37.0 | −1.0 | 37.0 | −1.0 | 38.3 | 37.4 |
| Naphthalene | 0.688 | 3.125 | 1.500 | 61 | 63.9 | −2.9 | 64.0 | −3.0 | 62.8 | 64.4 |
| Anthracene | 1.000 | 4.750 | 3.000 | 83 | 90.6 | −7.6 | 91.1 | −8.1 | 87.0 | 91.3 |
| Phenantrene | 1.000 | 4.750 | 1.063 | 91 | 90.6 | 0.4 | 90.5 | 0.5 | 87.0 | 91.3 |
| Styrene | 0.375 | 1.625 | 0.625 | 38 | 35.7 | 2.3 | 35.6 | 2.4 | 35.8 | 36.0 |
| Stilbene | 0.750 | 3.250 | 1.250 | 74 | 71.3 | 2.7 | 71.1 | 2.9 | 71.6 | 71.9 |
| Biphenyl | 0.750 | 3.250 | 1.250 | 71 | 71.3 | −0.3 | 71.3 | −0.3 | 71.6 | 71.9 |
| Butadiene | 0.000 | 0.000 | 0.000 | 3.5 | 0.0 | 3.5 | 0.0 | 3.5 | 0.0 | 0.0 |
| Fluorene | 1.181 | 8.444 | 9.111 | 76 | 76.1 | −0.1 | 76.6 | −0.6 | 45.5 | 76.0 |
| 1,3,5-Triphenyl benzene | 1.500 | 6.750 | 3.750 | 149 | 139.9 | 9.1 | 137.5 | 11.5 | 138.1 | 141.1 |
| Toluene | 0.375 | 1.625 | 0.500 | 35 | 35.7 | −0.7 | 35.7 | −0.7 | 35.8 | 36.0 |
| o-Xylene | 0.375 | 1.750 | 1.063 | 35 | 34.3 | 0.7 | 34.3 | 0.7 | 33.3 | 34.6 |
| Diphenyl methane | 0.750 | 3.250 | 1.250 | 67 | 71.3 | −4.3 | 71.6 | −4.6 | 71.6 | 71.9 |
| Naphthacene | 1.313 | 6.375 | 4.500 | 110 | 117.4 | −7.4 | 118.4 | −8.4 | 111.5 | 118.3 |
| Crysene | 1.313 | 6.375 | 4.625 | 116.5 | 117.4 | −0.9 | 117.5 | −1.0 | 111.5 | 118.3 |
| Pyrene | 1.188 | 6.250 | 4.500 | 108.9 | 101.0 | 7.9 | 100.1 | 8.8 | 91.1 | 101.6 |
| Perylene | 1.500 | 7.875 | 6.125 | 126.3 | 127.7 | −1.4 | 127.9 | −1.6 | 115.4 | 128.5 |

[a]Data obtained from Reference 10.
[b]Predicted from Equation 10.
[c]Values obtained by cross-validation.
[d]Predicted from Equation 12.
[e]Predicted from Equation 13.

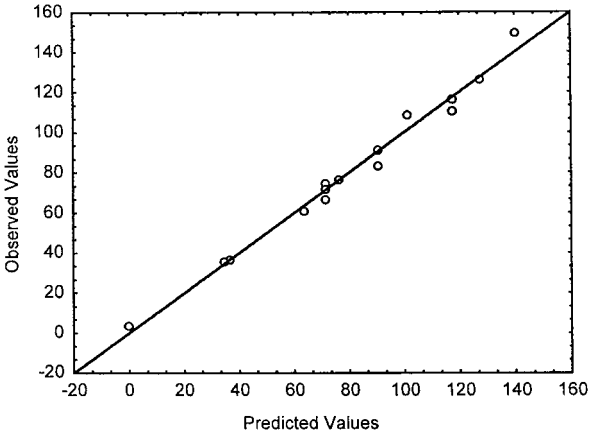


*Figure 2.* Experimental values versus predicted values from Equation 10 for the resonance energy of the compounds studied.

Here, $p_1$ and $p_2$ are the topological distances of the elemental paths that electrons 1 and 2 follow to reach point B. But, according to the de Broglie hypothesis, the wavelength associated with each electron is: $\lambda = h/(m \cdot v)$, where $h$ is the Planck constant and $m$ the electron rest mass. Thus, substituting in Equation 3 this leads to:

$$p_1 - p_2 = \frac{h}{m \cdot v \cdot d} = \Delta p_{12} \tag{4}$$

From here it leads that the kinetic energy required by the electron to reach a constructive interference is:

$$E_k = 1/2mv^2 = \frac{h^2}{2\,m\,d^2\,(\Delta p_{12})^2} \tag{5}$$

If we add the contributions from the difference of path lengths for which one of the paths is always of length one, then the following expression is obtained

$$E = \sum_{i=1}^{N}\sum_{j=1}^{N} E_{ij} = \frac{h^2}{2\,m\,d^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\delta(1,\ D_{ij})}{(\Delta p_{ij})^2} \tag{6}$$

$E$ is the total kinetic energy required for a constructive interference of all the π electrons; $N$ is the number of vertices; δ is the Kronecker symbol; and $D_{ij}$ is the topological distance i-j.

By analogy, we generalize the definition to any graph representing an aromatic system, and to any length of paths. This extrapolation pursues to simplify the computation of the new indices. A DP index of order K is defined as:

$$DP_k = \sum_{i=1}^{N}\sum_{j=i}^{N}\frac{\delta(K,\ D_{ij})}{(\Delta p_{ij})^2} \tag{7}$$

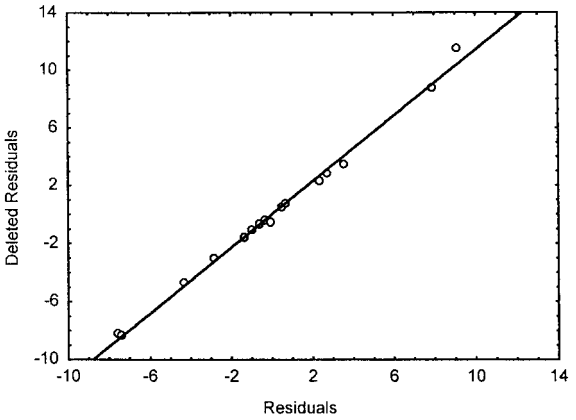


*Figure 3.* Comparison between the residuals obtained from Equation 10 and the deleted residuals in the cross-validation study for each compound.

Also, the valence DP indices, $DP^v$, are defined as:

$$DP_k^v = \sum_{i=1}^{N} \sum_{j=i}^{N} \frac{\delta(K,\ D_{ij})}{(\Delta p_{ij})^2}(1 + \Delta(EN_{ij})) \qquad (8)$$

where $\Delta(EN_{ij})$ are the differences of electronegativities between the i and j vertices placed at a topological distance K, considering pondered values of Pauling electronegativity taking the chlorine value as 2.

According to this definition, the DP indices would determine the values of the kinetic energy necessary for constructive interference of the electrons moving along the molecule in a diffraction experiment. In the last analysis, the electron mobility within the molecule would depend on the DP values.

## Results and discussion

In the study of the multilinear regression achieved with the resonance energy, RE, and DPs, the descriptor selection was carried out through factorial analysis. In fact, when the DPs were analyzed in this way, one factor was extracted explaining 92.6% of the variance and whose higher factor loading from the principal component analysis corresponded to $DP_1$, $DP_2$ and $DP_3$ (with 0.9518, 0.9976 and 0.9367, respectively).

Table 1 illustrates the values of DP indices as well as the experimental values of RE, for the selected hydrocarbons. It is to be noted that the third order index, $DP_3$, resolves the redundant values between pairs of compounds (see for example anthracene and phenantrene, or naphtacene and crysene), although in other cases, such as stilbene and biphenyl, there are isospectral values.

The regression equations obtained with one, two and three variables are:

$$RE = 83.97(\pm 5.13)\ DP_1$$
$$N = 17 \quad r = 0.9732 \quad SE = 9.12 \quad F = 1435 \qquad (9)$$

$$RE = 142.26(\pm 7.91)\ DP_1 - 10.89(\pm 1.56)\ DP_2$$
$$N = 17 \quad r = 0.9987 \quad SE = 4.57 \quad F = 2874 \qquad (10)$$
$$SEcv = 4.91$$

$$RE = 135.97(\pm 18.22)\ DP_1 - 9.05(\pm 5.01)\ DP_2 - 0.86(\pm 2.23)\ DP_3 \qquad (11)$$
$$N = 17 \quad r = 0.9987 \quad SE = 4.71 \quad F = 1807$$

These equations were obtained without the intercept, given the high standard error associated, which made it not significant statistically. The predictive ability by using only the index $DP_1$, Equation 9, is likely ($r = 0.9732$; $SE = 9.12$). The addition of a second variable, $DP_2$, Equation 10, allows a clear increase of F and, more interesting, lowers the SE value up to 50% (from $SE = 9.12$ to $SE = 4.57$). The presence of $DP_3$, Equation 11, is not significant statistically, so that Equation 10 was selected as the best predictive model.

Table 1 illustrates the calculated values of RE from Equation 10, as well as residuals for each compound. Moreover, we carried out a cross-validation study for the selected equation. The procedure involves the stepwise deletion of each compound and its value prediction from the correlation obtained from the $N - 1$ remaining data. This process is reiterated for each compound until all of them have been deleted one time. By means of the residuals it is possible to determine the standard error for the cross-validation analysis ($SE_{cv}$). The right positioned columns show the cross-validated values (subscript CV) for each one of the hydrocarbons. Furthermore, Figures 2 and 3 show the plot of observed versus calculated values as well as the deleted residuals versus residuals, for the results obtained from Equation 10. The excellent adjustment of the proposed model to the experimental results is noteworthy.

Equation 12 was obtained by using the indices up to the second order ones, i.e. $DP_1$ and $DP_2$, but considering as the training set the first seven hydrocarbons and as the validation set the rest of them (including butadiene). In contrast, Equation 13 was obtained by using the last 10 compounds as the training set and the first seven (including butadiene) as the validation set.

*Table 2.* Obtained classification matrix from the linear discriminant analysis study of a group of antibacterial compounds

| Group | Percent correct | Number of cases classified into group | |
|---|---|---|---|
| | | Active | Inactive |
| $DF1 = 0.089\ DP_4^V - 2.674$ | | | |
| Active | 71.6 | 116 | 46 |
| Inactive | 79.1 | 19 | 72 |
| Test active | 79.4 | 54 | 14 |
| Test inactive | 80.0 | 7 | 28 |
| $DF2 = 0.934\Delta^0\chi + 5.993\ DP_1^v - 3.635$ | | | |
| Active | 82.0 | 132 | 29 |
| Inactive | 90.1 | 9 | 82 |
| Test active | 88.2 | 60 | 8 |
| Test inactive | 91.4 | 3 | 32 |

The resulting equations are:

$$RE = 183.04(\pm 34.95)\ DP_1 - 20.21(\pm 7.65)\ DP_2 \quad (12)$$
$$N = 8 \quad r = 0.9990 \quad SE = 3.26 \quad F = 1510$$

$$RE = 144.48(\pm 10.71)\ DP_1 - 11.20(\pm 2.05)\ DP_2 \quad (13)$$
$$N = 10 \quad r = 0.9987 \quad SE = 5.58 \quad F = 1418$$

The two last columns in Table 1 show the prediction results obtained in this study.

It may be verified that a good prediction is obtained, excepting two cross-validated values corresponding to fluorene and pyrene, which show residuals higher than 15. Nevertheless, in most cases, the validated values demonstrate clearly the very good ability of the DP indices to predict RE. This was to be expected since, as pointed out above, this magnitude is closely related to the wavelength values, and therefore to the kinetic energy of the electrons moving into the aromatic ring as to reach a constructive interference.

Furthermore, since electron mobility clearly conditions other molecular properties, such as polarizability, on which, in turn, many biological and pharmacological characteristics depend, the possibility should be considered for the DP indices to show predictive content on this sort of properties. As illustration we have used a combined function including connectivity indices as well as DP ones, to discriminate the antibacterial activity by linear discriminant analysis (LDA). The objective of this technique, which is considered as one of the *pattern recognition methods*, is to find a linear function able to discriminate between two different classes of objects. The analysis is carried out using two large sets of compounds: one with proven pharmacological activity (in our case, antibacterial drugs), and the other with inactive compounds. The discriminant ability is tested by the percentage of correct classifications in each group; this is especially useful when the tested active–inactive compounds are not used as database. This is called a *cross-validation* test, in which topological indices act as independent variables discriminating activities.

The election of a discriminant function was carried out using the BMDP 7M package [11]. The method used for the selection of the descriptors was the F-Snedecor, and the classification criterion was the shortest Mahalanobis distance (distance of each case to the mean of all cases used in the regression equation).

For the LDA study a set of 193 structurally heterogeneous compounds with either antibacterial or non-antibacterial activity has been analyzed. Each group was separated in two: training and test groups. In this way, the discriminant function obtained can be validated.

The discriminant functions chosen were:

$$DF_1 = 0.089\ DP_4^v - 2.674 \quad (14)$$
$$N = 253 \quad F = 136 \quad U(\text{wilk's}) = 0.649$$

$$DF_2 = 0.934\ \Delta^0\chi + 5.993\ DP_4^v - 3.635 \quad (15)$$
$$N = 355 \quad F = 98 \quad U(\text{wilk's}) = 0.56$$

where $\Delta^0\chi$ is the difference between the valence and non-valence zero order connectivity indices. Such differences take into account the $\pi$ and lone pair electron properties [12, 13].

Table 2 summarizes the classification for the LDA matrix. By using only one index, namely $DP_4^v$, an overall accuracy of about 80% is obtained for both, training and test sets. However, if a second index is included, namely $\Delta^0\chi$, the overall accuracy increases to over 85%. The selection criterion was that a given compound is selected as antibacterial only if $DF_2 > 0$.

Tables 3 and 4 show the activity classification for a wide set of compounds used in the LDA study. It can be realized that both, training and test groups, get an average measure of correct prediction higher than 85%. On the other hand, in most cases, we work within

*Table 3.* Classification of compounds in the training series by the discriminant function obtained with Equation 15

| Compound | Prob (act.) | Class | Compound | Prob (inact.) | Class |
|---|---|---|---|---|---|
| Training active group | | | | | |
| Acedapsone | 0.248 | – | Flumequine | 0.435 | – |
| Acetosulfone | 0.534 | + | Guamecycline | 0.971 | + |
| Amifloxacine | 0.708 | + | Lenampicine | 0.970 | + |
| Apramycin | 0.993 | + | Mafenide | 0.146 | – |
| Azidamfenicol | 0.411 | – | Miloxacine | 0.809 | + |
| Azthreonam | 0.968 | + | Moxalactam | 0.952 | + |
| Butirosin | 0.994 | + | Nalidixic acid. | 0.296 | – |
| Carindacillin | 0.858 | + | Nifuradene | 0.774 | + |
| Cefazolin | 0.916 | + | Nifurpirinol | 0.657 | + |
| Ceforanide | 0.920 | + | Norfloxacine | 0.682 | + |
| Cefotiam | 0.942 | + | Ofloxacine | 0.848 | + |
| Cefoxitin | 0.891 | + | Pefloxacine | 0.667 | + |
| Ceftazidime | 0.983 | + | Pipacycline | 0.955 | + |
| Ceftezole | 0.916 | + | Pipemacine | 0.693 | + |
| Cephacetrile | 0.796 | + | Rifamide | 0.969 | + |
| Cephaloridine | 0.906 | + | Salazosulfadimidine | 0.708 | + |
| Cephradine | 0.746 | + | Streptomycin | 0.995 | + |
| Chloramphenicol | 0.250 | – | Sulbenicillin | 0.923 | + |
| Cyclacillin | 0.832 | + | Sumoxole | 0.507 | + |
| Dapsone | 0.220 | – | Talampicillin | 0.961 | + |
| Diathymosulfone | 0.559 | + | Tetroxoprim | 0.404 | – |
| Difloxacin | 0.844 | + | Ticarcillin | 0.932 | + |
| Doxycycline | 0.807 | + | Tobramine | 0.666 | + |
| Fleroxacine | 0.850 | + | Clinafloxacin | 0.975 | + |
| Training inactive group | | | | | |
| Acifran | 0.694 | – | Formothion | 0.922 | – |
| Acipimox | 0.741 | – | Genite | 0.906 | – |
| Acronine | 0.610 | – | Benzoic acid | 0.945 | – |
| Aldicarb | 0.954 | – | Blue acid | 0.153 | + |
| Allicin | 0.969 | – | Buspirone | 0.451 | + |
| Altretamine | 0.915 | – | Canthaxanthin | 0.935 | – |
| Amefenamine | 0.908 | – | Chlormezanone | 0.790 | – |
| Amitraz | 0.949 | – | Alprazolam | 0.746 | – |
| Antipirine | 0.841 | – | Erythrocentaurin | 0.878 | – |
| Antrafen | 0.195 | + | Etifoxine | 0.879 | – |
| Azapicyl | 0.865 | – | Fluoresone | 0.851 | – |
| Azaserine | 0.843 | – | Glutamic acid | 0.839 | – |
| Benceno | 0.974 | – | Hydroxyphenamate | 0.897 | – |
| Bufexam. | 0.897 | – | Hydroxyzine | 0.823 | – |
| Carprofen | 0.770 | – | Mephenoxalone | 0.680 | – |
| Ciclamate | 0.860 | – | Lovastatin | 0.683 | – |
| Clofibrate | 0.930 | – | Malaoxon | 0.868 | – |
| Clofibric acid | 0.913 | – | Malation | 0.884 | – |
| Dialifor | 0.802 | – | Mecarbam | 0.900 | – |
| Diclofen | 0.934 | – | Mitomycin | 0.130 | + |
| Dideoxycytidine | 0.401 | + | Nimustin | 0.755 | – |
| Diflunisal | 0.620 | – | Oxipizone | 0.620 | – |
| Doxorubine | 0.036 | + | Paracetamol | 0.906 | – |
| EDTA | 0.552 | – | Piroxicam | 0.487 | + |
| Feprazone | 0.812 | – | | | |

*Table 4.* Classification of compounds in the test series by the discriminant function obtained through Equation 15

| Compound | Prob (act.) | Class | Compound | Prob (inact.) | Class |
|---|---|---|---|---|---|
| Test active group | | | | | |
| 2 Gentamicine | 0.923 | + | Furaltadone | 0.955 | + |
| Amdinocillin | 0.779 | + | Gentacin | 0.907 | + |
| Amikacin | 0.994 | + | Glucosulfone | 0.998 | + |
| Amoxicillin | 0.896 | + | Isepamicin | 0.988 | + |
| Ampicillin | 0.807 | + | Methacycline | 0.807 | + |
| Arbekacin | 0.985 | + | Netilmicin | 0.907 | + |
| Azlocillin | 0.965 | + | Nitrofurazone | 0.819 | + |
| Carumonam | 0.988 | + | Penamecillin | 0.795 | + |
| Cefaclor | 0.678 | + | Penethamate | 0.715 | + |
| Cefbuperazone | 0.973 | + | Bacampicillin | 0.738 | + |
| Cefixime | 0.964 | + | Phenethicillin | 0.82 | + |
| Cefmenoxime | 0.955 | + | Phthalysulfonazole | 0.523 | + |
| Cefmetazole | 0.685 | + | Pipemidic acid | 0.693 | + |
| Cefonicid | 0.961 | + | Piperacillin | 0.959 | + |
| Cefoperazone | 0.985 | + | Pivampicillin | 0.852 | + |
| Cefpimizole | 0.998 | + | Ribostamycin | 0.984 | + |
| Cefpodoxime | 0.967 | + | Rifamicin | 0.964 | + |
| Cefsulodin | 0.973 | + | Rifampin | 0.991 | + |
| Ceftibuten | 0.947 | + | Rolitetracycline | 0.932 | + |
| Cefuroxime | 0.978 | + | Rosoxacin | 0.355 | – |
| Cefuzonam | 0.958 | + | Sancycline | 0.671 | + |
| Cinoxacin | 0.751 | + | Sulfisoxazole | 0.519 | + |
| Ciprofloxacin | 0.974 | + | Sulfoxone | 0.583 | + |
| Clometocillin | 0.754 | + | Sulguano | 0.602 | + |
| Clomocycline | 0.851 | + | Tobramycin. | 0.976 | + |
| Cloxacillin | 0.947 | + | Tosufloxacin | 0.958 | + |
| Dihydrostreptomycin | 0.994 | + | Miloxacin | 0.809 | + |
| Fenbenicillin | 0.82 | + | 1Gentamicine | 0.907 | + |
| Florfenicol | 0.203 | – | | | |
| Test inactive group | | | | | |
| AAS | 0.889 | – | Clofibric acid | 0.913 | – |
| Aminopirine | 0.866 | – | Clorofene | 0.949 | – |
| Benorylate | 0.811 | – | Cryptoxanthin | 0.949 | – |
| Apazone | 0.607 | – | Alpidem | 0.742 | – |
| Azacosterol | 0.913 | – | Emylcamate | 0.936 | – |
| Azobenzene | 0.954 | – | Meprobamate | 0.849 | – |
| Bumadizon | 0.869 | – | Prazepam | 0.900 | – |
| Butibufen | 0.945 | – | Beclobrate | 0.930 | – |
| Carmustine | 0.954 | – | Melinamide | 0.959 | – |
| Chlornaphazine | 0.976 | – | Theofibrate | 0.348 | + |
| Difenpiramide | 0.911 | – | Mitotane | 0.982 | – |
| Flurbiprofen | 0.883 | – | Monocrotophos | 0.877 | – |
| Ftorafur | 0.438 | + | Nabam | 0.978 | – |
| Glaphenine | 0.640 | – | Nicomol | 0.066 | + |
| | | | Nicotine | 0.881 | – |
| | | | Paraoxon | 0.704 | – |

*Table 5.* Results of antibacterial activity tests on different microorganisms for a set of compounds selected through the DF2 discriminant function, Equation 15. The tests were carried out through the method of Lennette et al. [15]

| Compound | Prob(act) | Class (act) | *E. coli* | *S. aureus* | *Paeruginosa* | *P. mirabilis* |
|---|---|---|---|---|---|---|
| Mordant Brown-24 | 0.747 | + | ++ | ++ | ++ | ++ |
| Neohesperidin | 0.989 | + | ++ | ++ | ++ | ++ |
| Silymarin | 0.965 | + | − | ++ | ++ | ++ |
| Hesperetin | 0.623 | + | ++ | ++ | ++ | ++ |
| Carminic Acid | 0.978 | + | + | + | + | + |
| Tartrazine | 0.935 | + | | | | |
| Morin | 0.778 | + | + | ++ | ++ | ++ |
| Fraxin | 0.869 | + | + | + | + | + |
| Niflumic acid | 0.526 | + | + | ++ | ++ | + |

a success probability higher than 80% (see Prob. in Tables 3 and 4).

In order to check the ability of Equation 15 for detecting antibacterial activity, we use a large database set including about 15000 commercial compounds, obtained from the Merck Index (Merck & Co., Inc., Rahway, NJ, U.S.A., Eleventh Edition, 1989) and the Sigma Aldrich catalogue (last edition, 1998). After screening a group of candidates showing $DF_2 > 0.0$, a set of compounds was selected as potential antibacterials. Table 5 illustrates some of these selected compounds, together with the results of the antibacterial activity experimental test, carried out against different strains from both gram + and gram − germs. According to these results it is remarkable that at least five compounds are obtained, namely mordant brown 24, neohesperidine dihydrochalcone, silymarine, hesperetine, morine and niflumic acid, showing a significant activity against at least two types of germs. It is to be emphasized that no previous report on antibacterial activity was found in the literature for these compounds.

It is also to be underlined that most of these compounds can be considered as new *leads*, since no structural similarity to any other known antibacterial is found.

## Conclusions

The novel indices of difference of path lengths, DPs, are new topological descriptors showing two important advantages with respect to most of those described in the literature: First, they are fully deduced from a physical model. In fact, they are derived from the formalism of the interferences occurring in a molecule as a consequence of the well-known electronic diffraction phenomena. Second, they encode important electronic information since they are a measure of the kinetic energy of the $\pi$ electrons moving along the rings in aromatic compounds. As a consequence, they must be useful descriptors, encoding non-redundant topological (or graph theoretical) information as compared with most of the indices outlined in the literature and they should thus be able to predict important physicochemical and biological properties depending on electronic mobility. This is demonstrated by their efficiency in predicting the resonance energy in aromatic hydrocarbons, as well as the antibacterial activity in a wide set of compounds showing considerable structural heterogeneity.

## Acknowledgements

The authors are indebted to Dr. Estrada, Dr. Mishra and Dr. de Julian-Ortiz for their help in the formal aspects of the manuscript as well as to the Spanish Ministry of Science (CICYT) (Grant SAF96-0158-C02-02) and to the Autonomic Government of Valencia (Grant GV-3300/95). The authors also thank J. Llompart for the use of his owner software for indices calculation, as well as professors Uruburu and Garcia-Lopez, from the Department of Microbiology, University of Valencia, for their help in performing the microbiological assays.

## References


1. Klein, D.J. and Randic, M., J. Math. Chem., 12 (1993) 81.
2. Basak, S.C., Med. Sci. Res., 15 (1987) 605.
3. Kier, L.B. and Hall, L.H., Molecular Connectivity in Structure–Activity Analysis, Research Studies Press, Letchworth, U.K., 1986, p. 225.
4. Hall, L.H. and Kier, L.B., Quant. Struct.-Act. Relat., 9 (1990) 115.
5. Kier, L.B. and Hall, L.H., Pharm. Res., 7 (1990) 801.
6. Gálvez, J., Garcia, R., Salabert, M.T. and Soler, R., J. Chem. Inf. Comput. Sci., 34 (1994) 520.
7. Pastor, L., García-Domenech, R. and Gálvez, J., Bioorg. Med. Chem. Lett., 8 (1998) 2577.
8. Gálvez, J., Garcia, R., Julian-Ortiz, J.V. de and Soler, R., J. Chem. Inf. Comput. Sci., 35 (1995) 272.
9. Gálvez, J., Gómez-Lechón, M.J., García-Domenech, R. and Castell, V., Bioorg. Med. Chem. Lett., 6 (1996) 2301.
10. Atkins, P.W., Physical Chemistry, Third Edition, Oxford University Press, Oxford, U.K., 1991, pp. 666–667.
11. Dixon, W.J., Manual BMD ‘Biomedical Computer Programs’, University of California Press, Los Angeles, CA, 1990.
12. Kier, L.B. and Hall, L.H., Quant. Struct.-Act. Relat., 10 (1989) 134.
13. Kier, L.B. and Hall, L.H., Pharm. Res., 6 (1989) 497.
14. Klages, F., Chem. Ver., 82 (1949) 358 .
15. Lennette, E.H., Balows, A., Hausler, W.J. and Shadomy, H.J., Manual de Microbiología Médica. Panamericana, 4th Ed., 1987, pp. 1191–1223.