

Core Electron Binding Energy (CEBE) Shifts Applied to Structure Activity Relationship (SAR) Analysis of Neolignans

M. Cristina A. Costa and Yuji Takahata*

Instituto de Química, Universidade Estadual de Campinas, CP 6154, 13083-862 Campinas - SP, Brazil

O deslocamento da energia de ligação do elétron do carvão (Δ CEBE) e o CEBE de átomos de carbono, calculados com o método semi-empírico HAM/3, foram utilizados como descritores no estudo das relações estrutura-atividade (SAR) de seis neolignanas. Os resultados obtidos demonstraram a eficiência deste tipo de descritores, nas análises de SAR. Usando-se cinco valores selecionados de CEBE, dos carbonos presentes nos anéis fenílicos das neolignanas, foi possível classificá-las nas duas categorias, ativa e inativa, usando-se os métodos HCA, PCA, KNN e SIMCA.

Core electron binding energy shifts (Δ CEBE's) and CEBE of carbon atoms calculated with the semi empirical HAM/3 method were shown to serve as a useful descriptor for SAR analysis of six neolignans studied. Using five selected Δ CEBE's of carbon atoms in the two phenyl rings of the compounds, the compounds were well separated by HCA, PCA, KNN and SIMCA methods.

Keywords: Δ CEBE, descriptor, neolignans, HAM/3, SAR, PCA

Introduction

The choice of molecular descriptors is one of the most crucial parts in the work of SAR/QSAR. Many descriptors have been suggested and employed. Many of them are successful and well accepted. Some descriptors that can be calculated by quantum mechanical methods have been recognized as useful in QSAR. However there still remains room to search for alternative and/or better descriptors than those in use, especially those descriptors that can be evaluated theoretically. One of our objectives is to look for more useful descriptors than thus far used, employing mainly quantum mechanical and/or other theoretical and computational methods. Lindberg *et al.* had shown that core electron binding energy correlate linearly to the Hammett sigma constants (σ) in substituted benzene derivatives.¹ Let us consider electrophilic aromatic substitution at *para* position of mono substituted benzene Ph-X as an example. X in Ph-X is a substituent such as OH, NH₂, NO₂ etc. Linderberg *et al.*¹ demonstrated the validity of an equation similar to equation 1:

$$\text{CEBE}(\textit{para} \text{ Carbon of Ph-X}) - \text{CEBE}(\text{Carbon of Ph-H}) \approx \kappa \sigma_X^p \quad (1)$$

Left hand side of equation 1 is the difference between core electron binding energy (CEBE) of *para* carbon of Ph-X and CEBE of carbon atom of benzene Ph-H, which is the reference molecule. The left hand side of equation 1 is called CEBE shift or Δ CEBE. The right hand side of equation 1 is a product between a constant κ and a Hammett sigma constant at *para* position, σ_X^p , of Ph-X. Equation 1 is an approximate equation. There are equations similar to equation 1 at other carbon atoms in the molecule such as *ortho* and *meta* positions in Ph-X. Linear relationship between Δ CEBE and Hammett sigma constants is not limited to monosubstituted benzenes. There are corresponding equations for multiply substituted benzenes. Straight lines were obtained by plotting experimentally observed CEBE values of Ph-X with respect to corresponding Hammett sigma constants σ_X .¹ This is a demonstration of the validity of equation 1. Recently we reconfirmed this² by calculating accurate Δ CEBE of the ring carbon in mono substituted benzene (Ph-X) in relation to the ring carbon in Ph-H. Density functional theory (DFT) was employed for the calculation of the Δ CEBE. Good agreement between the calculated CEBE and the Hammett σ constant³ of the corresponding atom was obtained.² Since Hammett sigma constant is one of the important descriptors in QSAR analysis,⁴ we can expect that Δ CEBE calculated theoretically can be also a useful descriptor in QSAR. Hammett sigma constants are

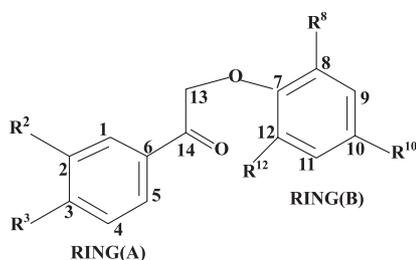
* e-mail: taka@iqm.unicamp.br

Table 1. Six studied neolignans. See Figure 1 for the positions of substituents “R” listed in the first line of the table

Compound	R ²	R ³	R ⁸	R ¹⁰	R ¹²	Activity
1	H	H	H	H	H	Inactive
2	H	H	H	-OCH ₃	H	Inactive
3	H	H	-OCH ₃	H	H	Inactive
4	-OCH ₃	-OCH ₃	H	-OCH ₃	H	Active
5	-OCH ₃	-OCH ₃	-OCH ₃	H	-OCH ₃	Active
6	-OCH ₃	-OCH ₃	H	-CH=CH-CH ₃	-OCH ₃	Active

usually determined experimentally. However in many drug molecules, Hammett sigma constant is not available. The object of the present work is firstly, to calculate Δ CEBE of a set of selected molecules whose Hammett sigma constants are not known, and secondly, to investigate whether or not the Δ CEBE is related to the biological activity of the molecules.

We chose six neolignans in which three of them are inactive and the other three are active against *leishmaniasis*. Figure 1 shows a skeleton of the neolignans and Table 1 list the six selected molecules and classes of biological response (active or inactive). They were taken from our previous publication.⁵ All the six molecules have the common basic skeleton.

**Figure 1.** Numbering system of neolignan.

Method of Calculation

We used the molecular geometry calculated by MM2 method previously.⁶ The semi-empirical HAM/3 (Hydrogenic Atoms in Molecules, version 3)⁷ method was used to calculate CEBE of the compounds. As far as we know, HAM/3 is the only semi-empirical method that can calculate CEBE's of a molecule. The widely used and well known semi-empirical method such as AM1 and ZINDO are not capable of calculating CEBE of a molecule. From our previous experience, average absolute deviation of CEBE's calculated by HAM/3 is expected to be about 1.50 eV.⁸ This is much larger than 0.3 eV that was attained by non-empirical DFT.² Only advantage of HAM/3 is its much higher speed of calculation in comparison to non-empirical DFT. Since neolignan is fairly large molecule

and calculation of CEBE has to be done one atom at a time, HAM/3 is a method of choice. We calculated CEBE's of 14 carbon atoms, C1-C14, in each molecule, that comprise of all the 12 carbon atoms in the two benzene rings plus two carbons that bridge the two benzene rings (see Figure 1). Δ CEBE's were calculated by equation 2, taking the difference between the calculated CEBE of each of the molecules and the value of 286.20 eV which is the CEBE of a carbon atom in an isolated benzene molecule calculated by HAM/3.

$$\Delta\text{CEBE} = \text{CEBE}(\text{molecule}) - 286.20 \quad (2)$$

Then, Fisher's weights of the Δ CEBE's were calculated. Some top greatest values of the weights were selected as useful descriptors for SAR analysis. Pattern recognition methods⁹ such as principal component analysis (PCA), hierarchical clustering analysis (HCA), K-nearest neighbors (KNN) and SIMCA were employed to study relation between the selected descriptors and the biological activity (SAR). The data were preprocessed by the method of autoscaling, then they were employed in the pattern recognition methods.

Results and Discussions

Table 2 and Table 3 list calculated CEBE's and Δ CEBE's respectively of five carbon atoms C1, C2, C4, C9 and C11 (See Figure 1) selected. These carbon atoms have the top five greatest Fishers' weights among the whole set of the 14 carbon atoms. The first three atoms, C1, C2, C4, belong to the A-Ring; the other two, C9 and C11, belong to the B-Ring of neolignan. Figure 2 shows a dendrogram produced with HCA using the Δ CEBE's in Table 3. The linkage method used was that of single link. The scale numbers on top part of the figure indicate similarity. The active compounds (**5**, **4** and **6**) are grouped together upper part of the figure, while inactive ones (**1**, **2** and **3**) are grouped together lower part of the figure. The two groups are well separated. Figure 3 shows score plot produced by PCA using the Δ CEBE's in Table 3. The x-axis represents

Table 2. CEBE's (eV), calculated with HAM/3, of the five carbon atoms C1, C2, C4, C9 and C11 (Figure 1) for the six compounds studied

Compound	C1	C2	C4	C9	C11
1	286.85	286.74	286.65	286.36	286.34
2	286.90	286.77	286.68	286.30	286.37
3	286.83	286.71	286.61	286.20	286.49
4	287.08	289.84	286.91	286.55	286.60
5	286.90	289.70	286.78	286.68	286.61
6	287.12	289.88	286.96	286.40	286.58

Table 3. Δ CEBE's (eV), calculated with equation 2 in the text, of the five carbon atoms C1, C2, C4, C9 and C11 (Figure 1) for the six compounds studied

Compound	C1	C2	C4	C9	C11
1	0.66	0.55	0.46	0.17	0.15
2	0.71	0.58	0.49	0.11	0.18
3	0.64	0.52	0.42	0.01	0.30
4	0.89	3.65	0.72	0.36	0.41
5	0.71	3.51	0.59	0.49	0.42
6	0.93	3.69	0.77	0.21	0.39

the first principal component (PC1), while y-axis represents the second principal component (PC2). All the three active compounds (**4**, **5**, **6**) are located extreme right hand side on x-axis, while all the three inactive compounds (**1**, **2**, **3**) are located extreme left hand side of x-axis. The active group is well separated from the inactive group. The two principal components (PC1 and PC2) are given in equations 3 and 4.

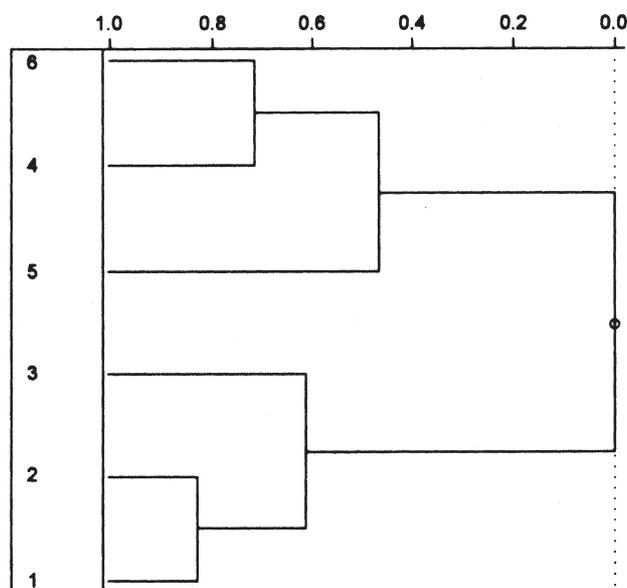


Figure 2. Dendrogram of the six neolignans resulted with HCA method. The numerical values on the scale on top part of the figure are similarity.

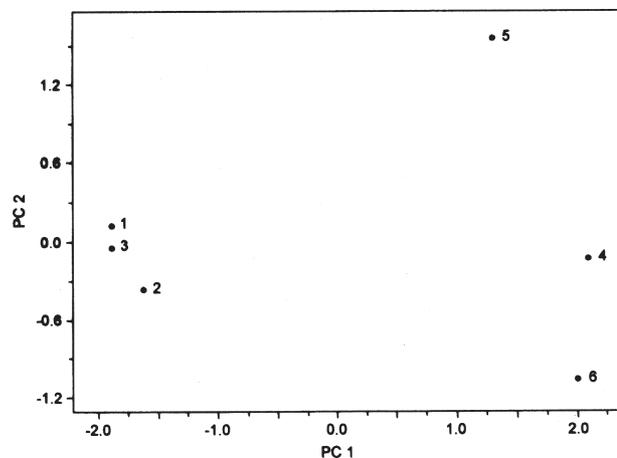


Figure 3. PCA score plot for the six neolignans.

$$PC1 = 0.43C1 + 0.50C2 + 0.48C4 + 0.38C9 + 0.44C11 \quad (3)$$

$$PC2 = -0.58C1 + 0.11C2 - 0.34C4 + 0.70C9 + 0.22C11 \quad (4)$$

PC1 explains 79.0% of variance and PC2 explains 15.0%. Cumulate variance up to PC2 is, therefore, 94.0%. Equation 3 indicates that all the five selected carbon atoms contribute more or less the same magnitude. The outstanding descriptors in equation 4 are C9 and C1. Figure 4 shows the loading graph for the five descriptors. The Δ CEBE's at C2 and C4 are mainly responsible for pulling the active group (**4**, **5** and **6**) towards right hand side in the score graphics. The Δ CEBE's of the three inactive compounds (**1**, **2** and **3**) are generally smaller than those of active compounds. This is especially true at C2 (Table 3). These are the reasons why the three inactive compounds (**1**, **2** and **3**) are located extreme left in Figure 3. We also used KNN, and SIMCA methods using the five selected

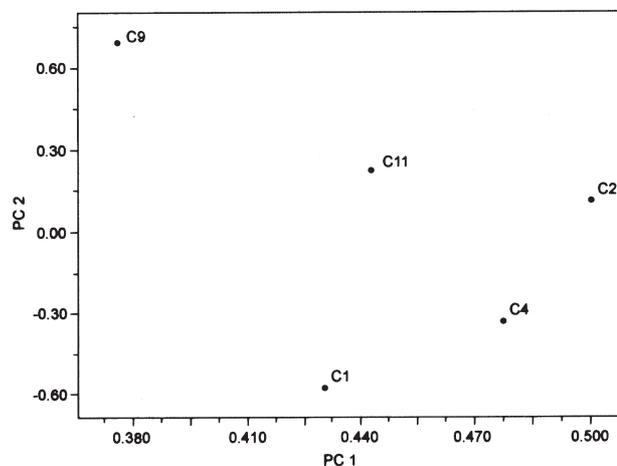


Figure 4. PCA loadings of the five selected carbons (Table 3).

descriptors. All of the 6 compounds were correctly classified by both KNN and SIMCA.

Instead of ΔCEBE (Table 3), we also used CEBE values themselves (Table 2) to see if they work as useful descriptors in SAR analysis with the pattern recognition methods. The results were completely identical to those obtained with ΔCEBE . This is due to the fact that the difference between CEBE and ΔCEBE is the constant (equations 1 and 2). After preprocessing of the data sets, the processed CEBE and ΔCEBE data sets become identical. Equation 1 can be rewritten in the form of equation 5,

$$\text{CEBE}(\text{Ph-X}) = \text{CEBE}(\text{Ph-H}) + \kappa\sigma_x \quad (5)$$

In equation 5, CEBE (Ph-H) is a constant because it is the CEBE of benzene which is the reference molecule. CEBE (Ph-X) is a linear function of variable σ_x with slope κ . The linearity of equation 5 was shown in figures in the literatures.^{1,2} Equation 5 shows that if ΔCEBE 's work as descriptors in SAR analysis, CEBE(Ph-X)'s themselves equally work as descriptors in SAR analysis. This situation is what we have confirmed numerically. We used mono substituted benzene (Ph-X) to discuss equations 1 and 5. But we can extend the discussions to multi substituted benzenes without loss of generality.

The Hammett equation, equation 6, correlates the equilibrium (or rate) constants (K) with the substituent constants σ_x for a system concerned X :

$$\text{Log}(K_x/K_0) = \rho\sigma_x \quad (6)$$

Here the subscript 0 denotes a reference system, and ρ , the reaction constant, is specific for the reaction considered. In case of chemical equilibrium under constant temperature, left hand side of equation 6 is linearly proportional to the difference of the change of free energy of Gibbs ($\Delta\Delta G$) in chemical/biological reactions between the system concerned and its reference system. $\Delta\Delta G$ is directly related to the relative affinity of interaction between the ligand and the biological target in the system concerned. This is the reason why Hammett sigma constants σ are so widely employed in the area where chemical and/or biological reactions are concerned. Comparison between equation 1 and equation 6 immediately reveals that ΔCEBE is a quantity that is linearly proportional to $\Delta\Delta G$. ΔCEBE has similar interpretability to the Hammett sigma constant σ .

Equation 5 indicates that CEBE (Ph-X) itself has a similar interpretability as ΔCEBE . Since in SAR studies, it is the *relative* quantity of $\Delta\Delta G$ that is important. Absolute value of $\Delta\Delta G$ is not necessary for the most of the

cases. Both ΔCEBE and CEBE are approximately proportional to $\Delta\Delta G$. This is the reason why they work in SAR analysis.

Number of compounds we worked in the present work is only six. This number is very small. The first reason why we worked with such a small set of molecules is that we wanted a quick and preliminary test if ΔCEBE (and CEBE) calculated with HAM/3 would serve as useful descriptor for SAR. Secondly, PCA works well even number of compounds are as small as six. This was demonstrated in our previous publication.¹⁰ We are currently working with a large number of compounds in order to see if ΔCEBE can really be one of useful descriptors in SAR/QSAR.

Conclusion

ΔCEBE (and CEBE) calculated with HAM/3 method was shown to serve as useful descriptor for SAR analysis of the six neolignans studied. Using five selected ΔCEBE 's, the compounds were well separated by HCA, PCA, KNN and SIMCA methods. CEBE and its shift (ΔCEBE) of an atom in a molecule reflect faithfully its chemical environment. Since ΔCEBE is linearly proportional to Hammett sigma constant, there is no surprise that ΔCEBE (and CEBE) demonstrated its usefulness in SAR. The conclusion thus far described is of a temporary nature, because the number of samples treated is very limited. Definite and general conclusion can be drawn only when a large number of samples with different types of molecules are treated.

Acknowledgements

MCAC thanks FAPESP for a post-doctoral fellowship and YT thanks CNPq for a research fellowship.

References

1. Linderberg, B.; Svensson, S.; Malmquist, P. A.; Basilier, E.; Gelius U.; Siegbahn, K.; *Chem. Phys. Lett.*, **1976**, *40*, 175.
2. Takahata, Y.; Chong, D.P.; *Bull. Chem. Soc. Jpn.*, **2000**, *73*, 2453.
3. Hammett, L. P.; *J. Am. Chem. Soc.*, **1937**, *59*, 96.
4. Kubinyi, H. In *QSAR: Hansch Analysis and Related Approaches*; Manhold, R., Krogsgaard-Larsen, Timmerman, T.; eds., VCH: Weinheim, 1993, vol. 1, p.4.
5. Costa, M.C.A.; Barata, L.E.; Takahata, Y.; *J. Mol. Struct. (THEOCHEM)*, **1995**, *340*, 185.
6. Costa, M.C.A.; Takahata, Y.; *J. Comput. Chem.*, **1997**, *18*, 712.
7. Åsbrink, L.; Fridh, C.; Lindholm, E.; *Chem. Phys. Lett.* **1977**, *52*: 63; *ibid*, **1977**, *52*, 69; *ibid*, **1977**, *52*, 72.

8. Takahata, Y.; *J. Mol. Struct. (THEOCHEM)* **1987**, *150*, 309.
9. Beebe, K. R.; Pell, R. J.; Seasholtz, M. B.; *Chemometrics: A Practical Guide*, John Wiley & Sons, Inc.: New York, 1998, pp. 56-182.
10. Vendrame R.; Ferreira, M. M. C.; Collins C. H.; Takahata Y.; *J. Mol. Graph. Model.* **2002**, *20*, 345.

Received: June 7, 2002

Published on the web: December 4, 2002

FAPESP helped in meeting the publication costs of this article.