We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Application of Self-Organizing Maps in Chemistry. The Case of Phenyl Cations

Daniele Dondi, Armando Buttafava and Angelo Albini
University of Pavia, Viale Taramelli 10-12, 27100 Pavia
Italy

1. Introduction

We have recently become interested in the application of Self-Organizing Maps (SOM) during a computational study on phenyl cations. As data accumulated, we realized that the analysis and interpretation of results, particularly when many variables were involved, could lead to a cognitive overload. In fact, it is for this reason that SOM found applications in chemistry in several problems, where the classification of large databases was not immediate, or the identification of the most characterizing properties of each class not obvious, since the a priori subdivision of the observed (complex) behaviour in more simple properties was not possible, as it is often the case.

The applications of SOM in chemistry are at present limited in number, but sufficient for indicating the potential of the method. The most important application is probably in the Quantitative Structure Activity Relationship (QSAR). The QSAR is a statistical method used in drug discovery where a correlation between biological activity (including desirable therapeutic effects and undesirable side effects) of chemicals (drugs/toxicants/environmental pollutants) with descriptors representative of molecular structure and/or properties is searched. Drug design has often the need to process enormous amounts of data, in which complex relationships have to be studied and modelled and is thus advantageously confronted by using SOM (Bienfait, 1994; Gramatica, 2007).

However, applications are really varied, including for example the analysis of complex mixtures such as raw oil spills (Fernàndez-Varela et al., 2010), the interpretation of spectra (Dow et al., 2004; Villman et al., 2008), studies of molecular conformation (Hyvönen et al., 2001), as well as the stucture of polymers (Llyod et al., 2008) or crystals (Willighagen et al., 2008), proteonomics (Herrero & Dopazo, 2002) and many others.

Closer to the topic of the present study, the SOM methods has been applied for recognizing the chemical properties of molecules, e.g. for assigning a degree of aromaticity (Alonso & Herradin, 2008), or more generally for predicting the chemical reactivity and its selectivity (Chen & Gasteiger, 1997; Noeske et al., 2006). As a matter of fact, phenyl cations had been all by unknown to chemists up to a decade ago, when it was discovered that electron-donating substituted phenyl halides, sulfonates and phosphates smoothly undergo heterolytic cleavage forming such intermediates in the triplet state, and that in this multiplicity these react efficiently with π, not with n, nucleophiles (Fagnoni & Albini, 2005). The synthetic potential of such intermediates seemed valuable, but the scope of both generation and

reaction required extensive work. Thus it was of interest to carry out computational studies and to apply the SOM method to the data case in order to recognize and predict the structural and chemical properties.

We used self-organizing maps, as developed by Kohonen (Kohonen, 2001; Wehrens et al., 2007). This is an unsupervised neural network that preserves on a two dimensional plane topological relations. We introduced as many as possible optimized geometric parameters and searched for structure similarities. Indeed a visually effective result was obtained through a Sammon map, which clearly separated triplet and singlet cations and helped to rationalize the different substituent effect on the two species.

In the following, the approach used will be illustrated and new examples will be discussed.

2. Description of the problem

Density Functional Theory (DFT) is a computational approach that has shown to be able to determine accurately energy differences, geometries and spectroscopic data of chemical compounds. The method is particularly useful if applied to transient species, such as excited states or reaction intermediates, for which a direct observation could be difficult.

During the last years, in our laboratories a reaction of photoarylation was studied (see Scheme 1).

$$Y = N(CH_3)_2, \ OH, \ OCH_3, \ SCH_3$$

$$X = Cl, \ F, \ PO_4Et_2, \ SO_3CH_3, \ SO_3 \ CF_3$$
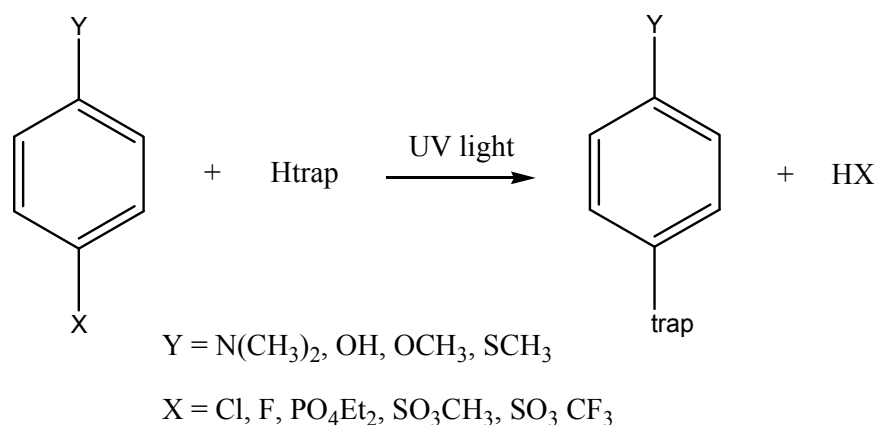
Scheme 1. Photoinduced arylation

This process is particularly interesting since such arylations can otherwise be obtained only under catalysis by transition metals, a highly versatile method that suffers however of some limitations, in particular related to the catalysts themselves, often delicate, expensive and polluting. Substituting light for heavy metal catalysis makes the reaction more compatible with the principles of green chemistry, i.e. a photoinduced reaction is more ecosustainable (Fagnoni & Albini, 2005).

We have demonstrated that the active chemical intermediate is the aryl cation, and more particularly the triplet spin state. Such an intermediate is highly reactive (it is a carbocation), but is nevertheless characterized by a high selectivity towards suitable traps. The other spin state, the singlet, is on the contrary completely unselective and reacts at diffusion controlled rate practically with everything. As a result, when generated in solution it typically reacts with the solvent.

The computational approach via DFT has been used often for cations and in the present case a first gratifying result was that the computed UV spectrum of the triplet cation was in

accordance with the experimentally determined spectrum of the intermediate, as obtained by flash photolysis (Manet et al., 2008).

From preliminary studies, it is known that the geometry of singlet and triplet phenyl cations differs significantly. In fact, the triplet phenyl cations have a more regular hexagonal geometry with respect to the singlet phenyl cation.

Parent phenyl cation

triplet   singlet
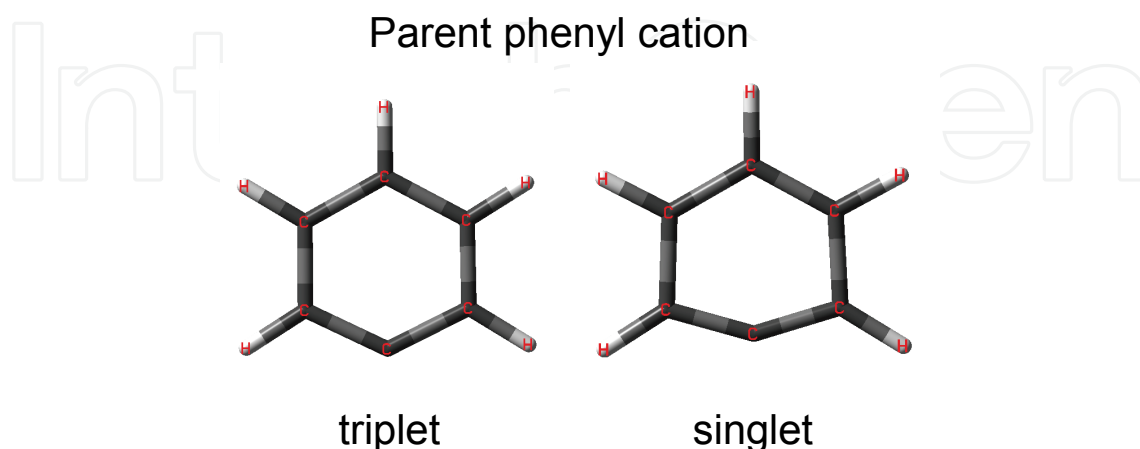
The introduction of substituents in such high-energy intermediates may induce important effects on structure and chemistry with respect to the parent cation. However, it is not obvious how this effect will operate. Singlet phenyl cations have a $\pi^6 n^0$ structure, that is the positive charge is localized at the divalent carbon and lies in the plane of the molecule. Therefore, mesomeric effects should not operate. On the other hand, triplets have a $\pi^5 n^1$, that is a diradicalic character, one radical site at the divalent carbon, one delocalized over the aromatic ring, which has thus a radical cation character. In this case, mesomeric effects may be significant. For simplicity, we considered monosubstituted derivatives, with the group in the three positions, ortho, meta and para, as indicated in the formulae below, where Y stands for an atom or a group of atoms.
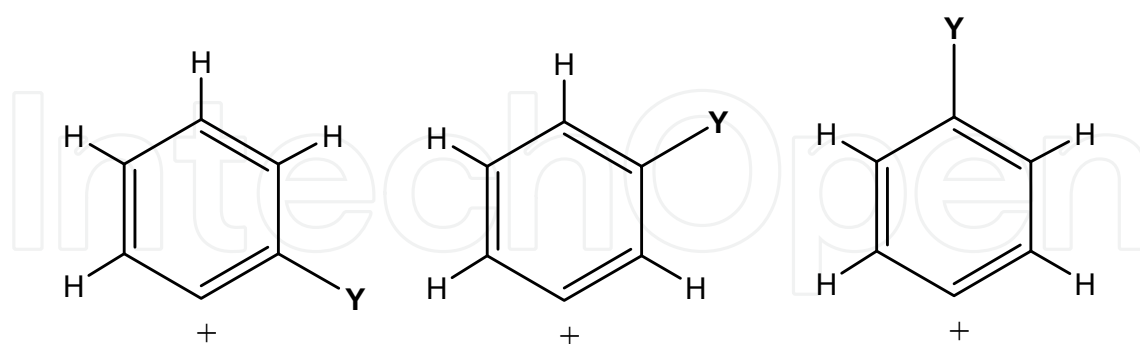
3. Data analysis through SOM

The target of the investigation was determining which effect had substituents (and their position) on structure and energy of phenyl cations and how these difference were reflected in the chemical reactivity. Perhaps unexpectedly, in view of the considerations above, the singlet cations had a geometry heavily deformated, with strong dependence on the substituent, despite the fact that these species conserved the intact aromatic sextet. On the

contrary triplets, which had lost the sextet, maintained quite closely the planar hexagonal geometry. The rationalization was not obvious. In order to have a significant results, a sufficiently differentiated choice of substituents had to be considered, avoiding on the other hand to overcome a manageable number. We thus decided to consider eight substituents Y (in the three positions), chosen among those expected to induce large changes due to the electron-donating or –abstracting effect, and to compare their properties with those of the parent cation (Lazzaroni et al, 2008, 2010). The groups, ordered from electron-donating to electron-withrawing were: NH_2, SCH_3, OCH_3, CH_3, $Si(CH_3)_3$, H, CN and NO_2.

Even in this oversimplified space, these corresponded to a total of 8 (substituents) x 3 (positions) x 2 (spin states) = 48 + 1 structures that were calculated. It was immediately apparent that the comparison by hand of 49 structures in a 3D space was scarcely promising due to the large number of data available. On the other hand, limiting the choice to a few key geometrical parameters was no appealing alternative, due to the lack of objectivity in the choice of such parameters.

This thus appeared as a suitable case for making recourse to unsupervised learning, which can help to find correlations and similarities among large series of data, if a correct input is given.

In the data set, we avoided the use of parameters directly related to the substituent Y, because a trivial categorization factor would be otherwise inserted. Thus, the length of the C-Y bond was not included because it depended on the substituent's nature). For this reason, only parameters related to the carbon skeleton of the aryl cation ring were introduced. In order to describe exhaustively the structure of the six carbon atoms in the space, bond lengths, bond angles and dihedral angles of the phenyl ring were introduced.
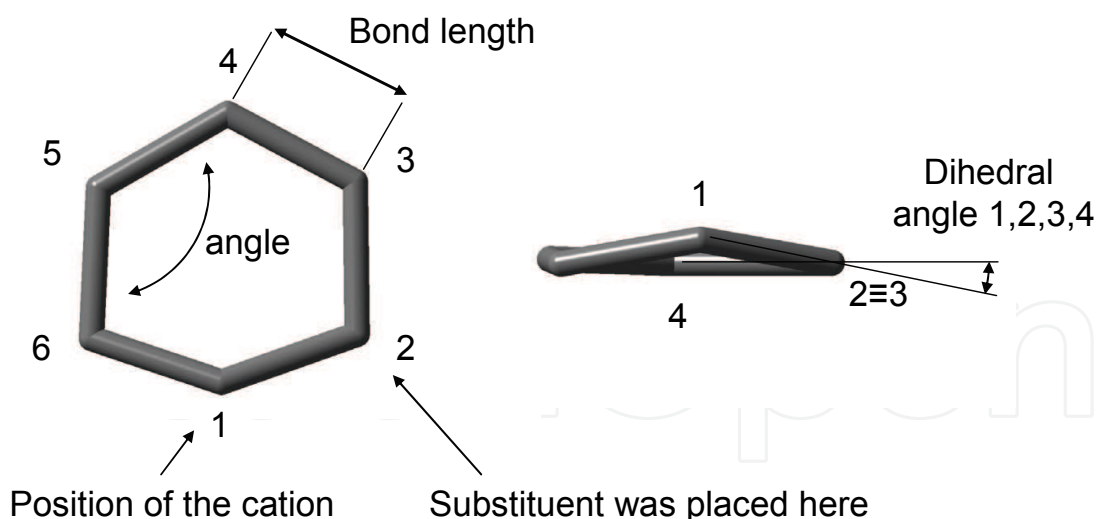

Fig. 1. Example of the geometrical data inserted as input

Furthermore, as the divalent carbon (C_1) underwent the most extensive geometrical modifications, all parameters were inserted starting from that carbon, and the shortest path to reach the substituent was adopted. A graphical representation of the choice of parameters and the numbering is given in Figure 1. This choice is essential for the correct application of the SOM.

The complete training set used is given in Table 1.

1.397	1.378	1.444	1.436	1.378	1.404	127.020	116.152	119.540	121.580	118.155	117.156	0.000	0.007
1.344	1.410	1.423	1.415	1.415	1.341	140.850	105.290	122.150	120.510	119.940	107.310	9.130	1.920
1.397	1.378	1.440	1.440	1.379	1.397	126.790	116.760	119.400	120.890	119.400	116.760	0.005	0.000
1.360	1.400	1.430	1.430	1.400	1.360	133.780	108.790	120.630	119.780	120.660	108.750	13.100	0.360
1.406	1.376	1.444	1.443	1.377	1.406	127.140	116.000	120.500	119.850	120.490	116.010	0.024	0.220
1.327	1.433	1.403	1.403	1.433	1.326	146.720	104.120	123.090	118.850	123.090	104.110	0.317	0.117
1.405	1.377	1.442	1.442	1.377	1.405	127.950	116.000	118.990	122.060	118.980	116.000	0.000	0.000
1.328	1.429	1.407	1.407	1.429	1.328	147.560	104.150	121.640	120.860	121.640	104.150	0.011	0.003
1.412	1.378	1.434	1.434	1.378	1.412	127.680	115.680	119.320	122.330	119.320	115.680	0.000	0.004
1.327	1.435	1.395	1.395	1.435	1.327	147.320	104.000	121.770	121.120	121.770	104.000	0.009	0.005
1.352	1.447	1.440	1.375	1.428	1.397	126.020	115.800	120.650	119.640	120.370	117.520	0.004	0.004
1.322	1.460	1.420	1.384	1.449	1.324	144.680	107.720	118.450	121.970	123.420	104.230	5.760	3.678
1.353	1.442	1.442	1.377	1.420	1.397	126.580	115.730	119.990	120.380	120.080	117.260	0.013	0.022
1.324	1.468	1.437	1.374	1.455	1.324	141.400	108.270	118.480	121.950	120.400	107.790	6.835	3.674
1.361	1.416	1.464	1.386	1.404	1.423	126.940	116.060	118.780	122.380	118.600	117.250	0.010	0.013
1.324	1.447	1.402	1.393	1.434	1.326	147.600	105.700	118.890	122.350	122.610	102.830	0.006	0.006
1.360	1.424	1.460	1.381	1.412	1.417	126.620	114.800	120.980	120.640	118.930	118.030	0.000	0.000
1.326	1.446	1.405	1.392	1.437	1.326	147.160	105.480	119.950	121.110	123.230	103.050	0.014	0.009
1.430	1.434	1.374	1.431	1.413	1.360	125.360	117.480	118.380	121.720	121.080	115.980	0.000	0.000
1.352	1.444	1.395	1.393	1.430	1.328	145.600	100.940	125.000	121.610	118.400	108.240	3.089	2.073
1.422	1.441	1.373	1.423	1.417	1.360	125.300	117.000	119.020	121.300	121.030	116.350	0.000	0.000
1.380	1.446	1.377	1.412	1.429	1.336	136.630	109.140	118.870	121.970	122.490	106.030	9.451	1.721
1.440	1.420	1.380	1.446	1.404	1.367	127.440	115.500	119.640	121.550	120.860	115.020	0.023	0.012
1.326	1.450	1.392	1.397	1.426	1.333	148.870	100.360	124.390	120.980	119.960	105.440	0.007	0.003
1.436	1.423	1.380	1.440	1.410	1.363	125.680	117.780	118.240	121.410	121.620	115.270	0.000	0.000
1.343	1.444	1.391	1.395	1.438	1.322	147.150	103.310	122.250	121.290	121.530	104.470	0.006	0.004
1.394	1.385	1.427	1.435	1.385	1.389	126.822	116.882	119.344	120.823	119.844	116.286	0.000	0.000
1.338	1.410	1.409	1.415	1.405	1.341	138.272	107.636	120.980	119.909	121.511	107.313	10.663	1.108
1.405	1.381	1.425	1.425	1.381	1.405	127.815	116.098	117.660	124.669	117.660	116.099	0.139	0.076
1.330	1.427	1.395	1.395	1.427	1.330	148.485	103.695	120.350	120.425	120.350	103.695	0.084	0.049
1.394	1.376	1.434	1.434	1.376	1.394	127.066	115.915	120.936	118.807	120.940	115.910	2.941	7.298
1.315	1.444	1.390	1.393	1.438	1.318	145.545	104.524	123.598	118.084	123.422	104.828	0.003	0.000
1.354	1.422	1.429	1.379	1.405	1.387	126.914	115.380	120.221	120.651	119.640	117.195	0.005	0.003
1.318	1.457	1.417	1.377	1.439	1.321	143.067	107.079	119.179	122.218	120.420	107.115	3.925	1.714
1.383	1.385	1.440	1.403	1.392	1.417	127.366	113.267	123.263	119.813	118.976	117.312	0.000	0.018
1.323	1.425	1.391	1.399	1.428	1.334	146.584	103.116	124.085	119.624	121.662	104.930	0.017	0.008
1.364	1.404	1.457	1.392	1.387	1.416	127.245	116.565	117.542	123.395	118.320	116.932	0.000	0.003
1.310	1.448	1.393	1.392	1.416	1.331	148.014	100.910	123.280	122.595	117.920	107.281	0.132	0.103
1.398	1.410	1.378	1.408	1.402	1.363	123.640	119.104	118.505	120.666	121.334	116.751	0.259	0.237
1.333	1.448	1.381	1.396	1.426	1.325	145.077	104.011	121.794	121.794	120.732	105.748	3.741	0.189
1.417	1.389	1.398	1.440	1.395	1.379	124.639	120.282	116.617	121.912	121.236	115.269	2.046	1.824
1.403	1.411	1.381	1.440	1.389	1.371	128.776	114.430	120.001	121.667	120.542	114.585	0.164	0.103
1.294	1.452	1.387	1.395	1.420	1.333	149.011	102.573	121.807	121.460	120.847	104.301	0.004	0.000

Table 1. Complete training set used. The first group of six columns are bond lengths (units are Ångström), the second group are angles (degrees) and the third dihedral angles (degrees). The last column is the coded name of the phenyl cation (see text for a description)

For the SOM optimization, we started with 1000 cycles and with an initial learning rate parameter (alpha) of 0.05. The map was then refined with 10000 cycles starting with alpha = 0.02. A bubble neighbouring was used and the radius was 10 for the first step and 3 for the last step.

The software package utilized for the calculations of SOMs was SOMPAK (Kohonen, 1997), a program developed by the Helsinki University of Technology.

A first trial with a rectangular 2x2 grid was carried out in order to verify the correctness of the input data; indeed, the database was correctly divided into two classes, the first one containing only singlets and the second one only triplets.

A more advanced level of information could be obtained by increasing the size of the map. The program used allowed the creation of two different lattice arrangements of neurons: rectangular and hexagonal. After some experimenting, we found convenient to adopt the rectangular lattice.

The calculated rectangular 5x5 map containing all the phenyl cation intermediates is reported below (Table 2). The numbering in the table is referred to the substituent position, as indicated in Figure 2, while the spin state is specified by the capital letter superscripts S for singlets and T for triplets. The formula of the substituent Y is placed after the number and H is used for the unsubstituted (parent) phenyl cation.
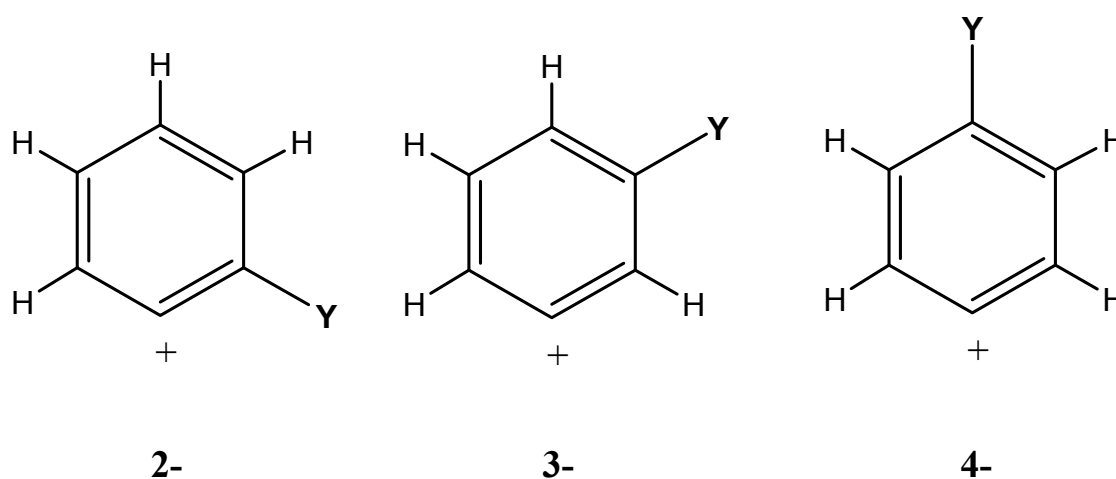
Fig. 2. Substituent numbering used in Table 2

The differences in geometry of singlet and triplet phenyl cations is well apparent, with the two classes lying in different positions in the table.

In this 5x5 map, 17 cathegories were represented, each one of them corresponding to an active neuron.

By analyzing the position of the entries in the table, some generalizations could be made; for example, when considering the singlets, substituents such as CN, NO_2, $Si(CH_3)_3$ and CH_3 were all grouped in the bottom left corner, close to the parent cation.

A chemical meaning can be immediately associated with this result. Thus, all of the substituents above are known to be electron-withdrawing or almost electron-neutral groups. The presence of an empty sp^2 orbital at C_1 in the place of a σ bond makes little difference for the intact π^6 system with which the substituents mainly interact, resulting in a moderate effect on the geometry of the phenyl cation.

On the other hand, cations bearing electron-donating substituents are found in the left part of the table, but their distribution is more spread across the table, with an effect that varies with the position in a different way for each substituent. The substituents with a strong mesomeric effect, such as amino, methoxy and thiomethoxy group cluster at the upper left corner. Clearly, the electron-rich π system does interact with the empty σ orbital, and both the nature and the position of ring substituents are important in determining the final geometric parameters.

As for the triplets, these are grouped in three big classes in the rightmost column. Each of them containing (mainly) cations bearing the substituent in the same position, regardless of their -donating or -withdrawing character. Apparently, with triplets the effect of the position of the substituent overcame the electronic effect in causing a deformation of the phenyl cation.

4-OCH$_3^S$ 4-NH$_2^S$ 2-NH$_2^S$ 4-SCH$_3^S$				4-Si(CH$_3$)$_3^T$
	2-SCH$_3^S$	3-NH$_2^S$ 3-SCH$_3^S$		3-NO$_2^T$
2-OCH$_3^S$	3-OCH$_3^S$		2-CH$_3^T$	4-CH$_3^T$ 3-OCH$_3^T$ 3-NH$_2^T$ 3-CN^T 3-SCH$_3^T$
2-CH$_3^S$ 3-Si(CH$_3$)$_3^S$		2-Si(CH$_3$)$_3^T$	4-OCH$_3^T$ 4-CN^T H^T 3-CH$_3^T$ 4-NO$_2^T$	4-NH$_2^T$ 4-SCH$_3^T$
4-CN^S 2-CN^S H^S 4-NO$_2^S$ 3-NO$_2^S$ 2-Si(CH$_3$)$_3^S$	4-CH$_3^S$ 3-CH$_3^S$ 3-CN^S 4-Si(CH$_3$)$_3^S$		3-Si(CH$_3$)$_3^T$	2-OCH$_3^T$ 2-NH$_2^T$ 2-CN^T 2-SCH$_3^T$ 2-NO$_2^T$

Table 2. 5x5 SOM map obtained for the complete set of calculated phenyl cations

Another useful information is given by the quantization of the error regarding each of the entries of the map. The quantization error is inversely proportional to the matching of each entry line (in our case each phenyl cation) within the assigned class.

Figure 3 shows the pictorial representation of the quantization errors for the 5x5 calculated map; each one of the grid intercepts has a corresponding element in Table 2. Circles with small areas are related to small quantization errors. As can be seen in the figure, a small number of neurons corresponds to classes having most of the best-matching entries. These neurons define a sharp-defined class, i.e. the entries located by this neuron are most similar among them. Considering those of the 'sharp' classes that contain more than one entry, it is found that these corresponds to the four classes evidenced in Table 2, one for the singlets with bearing no electron-donating substituent, three for the triplets with the substituent in each of the three position (that is the classes at positions 1;5, 4;4, 5;3 and 5;5).

Although the distribution in a table does give useful information about any alikeness, a more quantitative representation of class similarities in a calculated map is the projection of neurons through the Sammon's map (see Figure 4).

The label assigned to each point is that of the first entry in the class identified with that neuron. A look at Figure 4 makes it apparent that all of the triplets (squares) lie within a much smaller area with respect to that occupied by singlets (circles). Thus, even if the

number of classes belonging to triplets is similar to that of singlets, the former ones are subject to a smaller geometric variation.

Concerning singlets, the map shows how they are rather scattered on the Sammon's projection. The parent phenyl cation is located at coordinates -18, -10. Electron-donating substituents appear to cause a deformation that increases when passing from substituents in position 2 to the same in position 4.
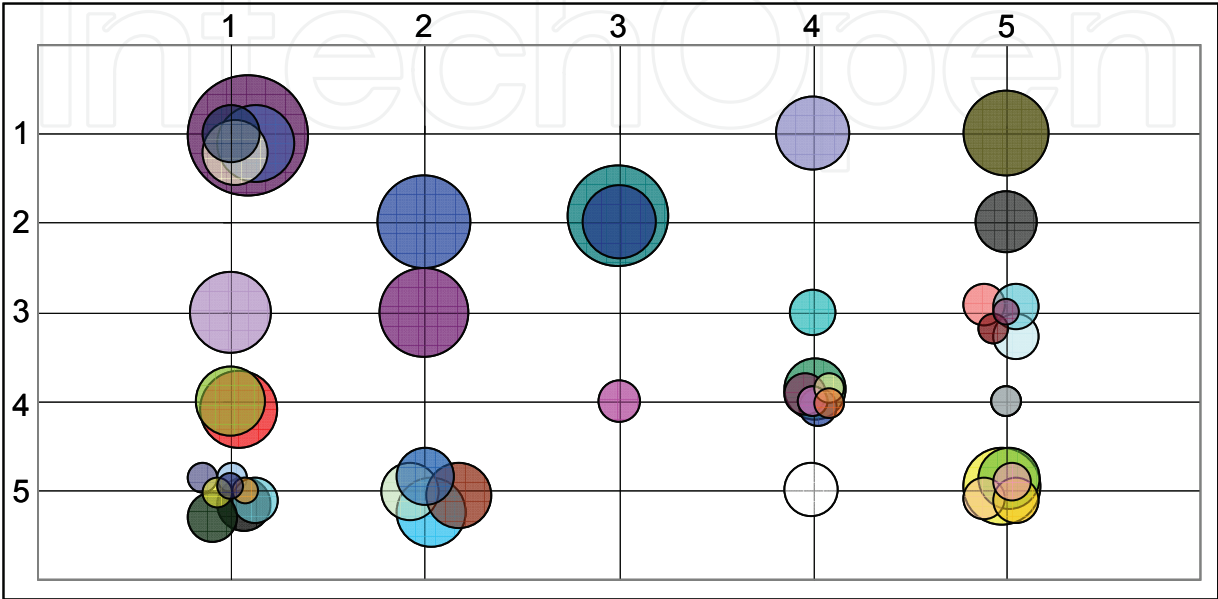

Fig. 3. Quantization error for the calculated 5x5 SOM. The area of the circle is proportional to the quantization error for that entry. Please note that overlap between circles is diminished because a random value was added to the position of the center
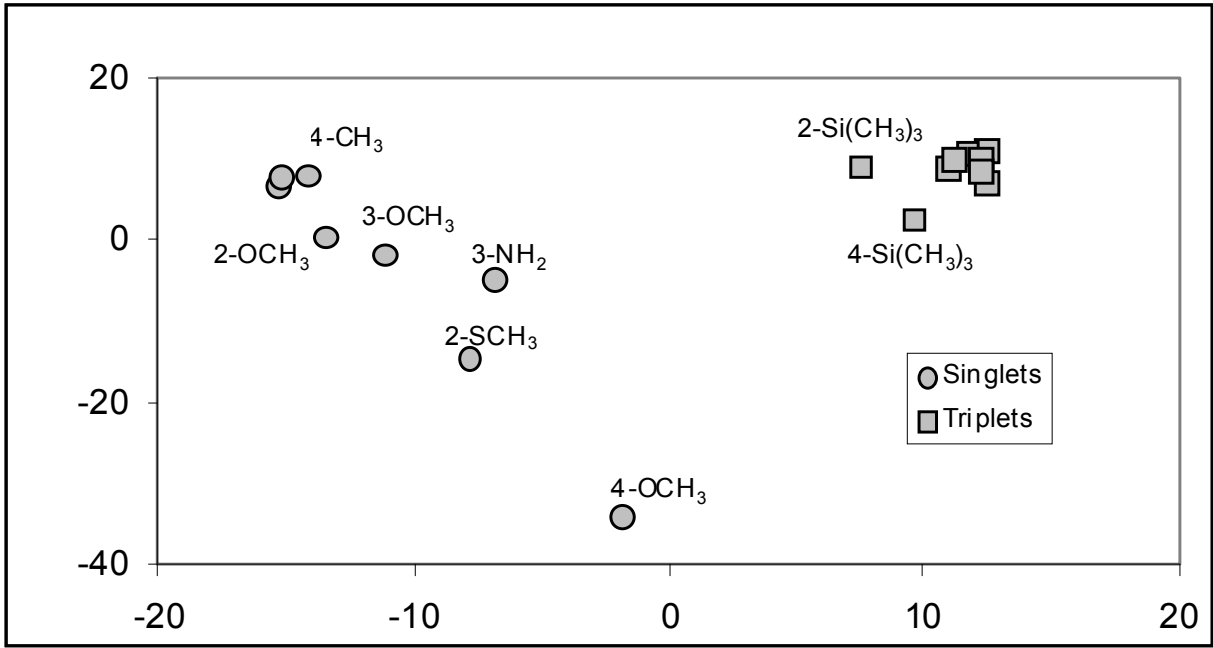
Fig. 4. Sammon's projection for the 5x5 map showed in table 1. Axes are in arbitral units

The above observation affords indications for assessing the effect of substituents on the geometry of phenyl cations. However, arriving at recognizing the key geometrical parameters for each class requires a different approach. This is possible by having recourse to the analysis of 'activation planes'. During the SOM optimization a plane of activation is created for each input variable. In our case we have three groups of six variables each, viz. bond lengths, bond angles and dihedral angles.
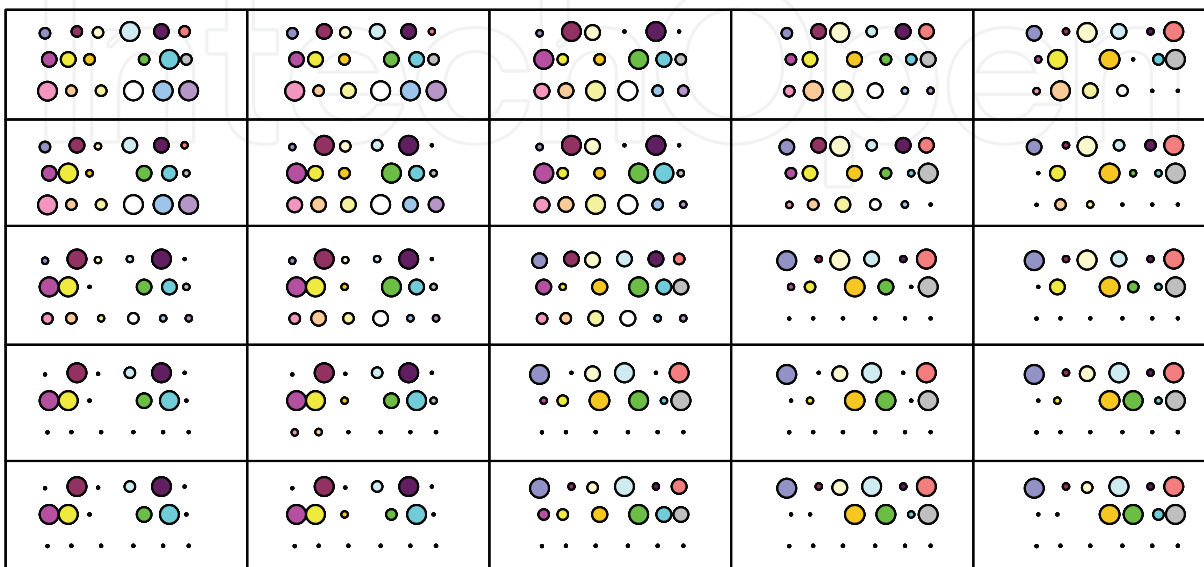

Fig. 5. Activation planes for each element of the grid

4-OCH₃ˢ   4-CNˢ
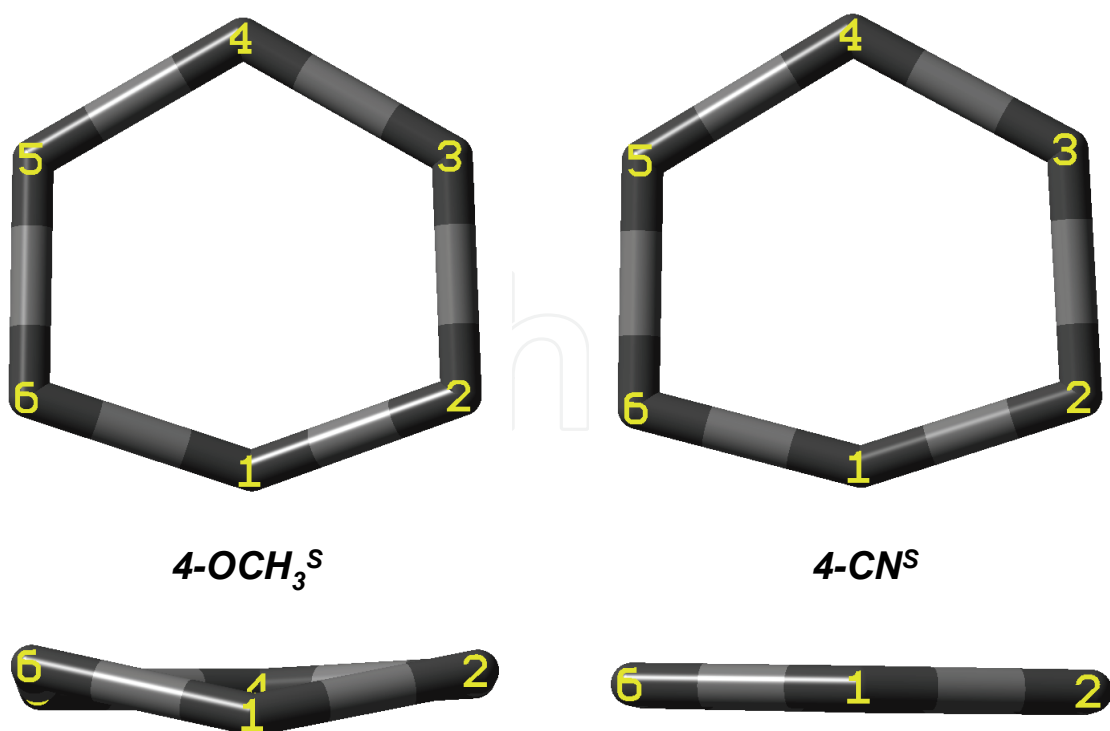
Fig. 6. Geometries of carbon rings for the phenyl cations 4-OCH₃ˢ and 4-CNˢ, top and side view

In Figure 5 these values are shown as circles having an area proportional to the weight of the corresponding input. For every element of Table 2 three lines (bond length, angle, dihedral angle) containing six circles each are a representation of the importance of the input data. As it can be seen in the figure, even table elements (and thus neurons) having no class members have nonzero values. In this picture, is also apparent the meaning of the word neighbourhood in this context (and thus the importance of rectangular or hexagonal grid): neurons in close contact have similar activation weights.

Let us consider the leftmost column of grid shown in Figure 5. It is apparent that when proceeding from the top to the bottom of the figure, the classification algorithm changes from one characterized by a high weight of the dihedral angles to one characterized by a high weight for specific bond lengths (second and fifth bond) and bond angles (1,2,4,5).

Notice that only few of these characteristics are visible by observing the actual geometries (see Figure 6).

In fact, only the dihedral angle differences are immediately appreciated in the side view, while the variation of bond lengths are better visible by looking at the corresponding rows in Table 2. The largest effect among the out-of-plane variations is the lengthening of bonds 2 and 5 for the CN-substituted derivative. This effect could be reasonably assigned to the stabilization ot the radical site by the CN substituent. Correspondingly, the methoxy group stabilizes the charge by the resonance electron donation (see the formulae in Figure 7).
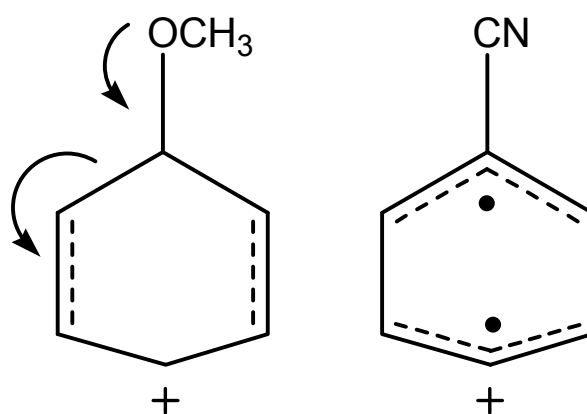

Fig. 7. Contribution of the electronic effect to the stabilization of singlet phenyl cations

In the same line, it may be asked which are the geometrical key parameter for the three distinct classes of triplets located at 4;4, 5;3 and 5;5. Figure 5 tell us that the dihedral angles do not contribute significantly (third row of six circles). In fact, all of these structures lie almost perfectly in a plane.

For these phenyl cations, the structures can be roughly represented as indicated in Figure 8, where the dashed double bond is slightly shorter with respect to other bonds. Since this effect, which is due to the electronic stabilization by the substituent, is quite small, almost no effect is visible observing the 3D structures (also in this case, a close scrutiny of the data in Table 2 reveals the effect, however).

Summing up, the various analyses carried out in this work allow a better understanding of the structure of phenyl cations and the effect of substituents on it. This is much more complex than that obtained by one- or two-parameters linear relationships (Lazzaroni et al., 2008, 2010) and not easily framed in the usual mesomeric/inductive effects. On the other hand, patterns are recognized and may have some predictive effect for further substituents

or for suggesting new chemistry (notice, as an example, the diradicalic nature of triplets, depending only from the position of the substituent, for which is not known at the moment any corresponding reaction.
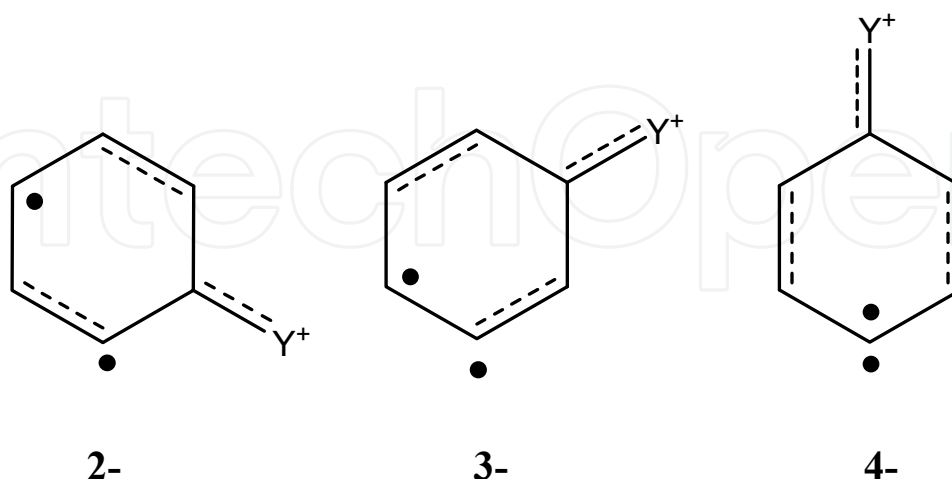

Fig. 8. Schematic structures for 2-, 3- and 4- substituted triplet phenyl cations

Finally, the important question of the relation between energy and geometry in these cations can be confronted. For this aim it is convenient to introduce a isodesmic chemical equation in which a hypothetical reaction is considered. One of these reactions has to be assumed as the reference zero, and in this case the unsubstituted phenyl cation was chosen. The isodesmic reaction considered is thus that shown in Figure 9.
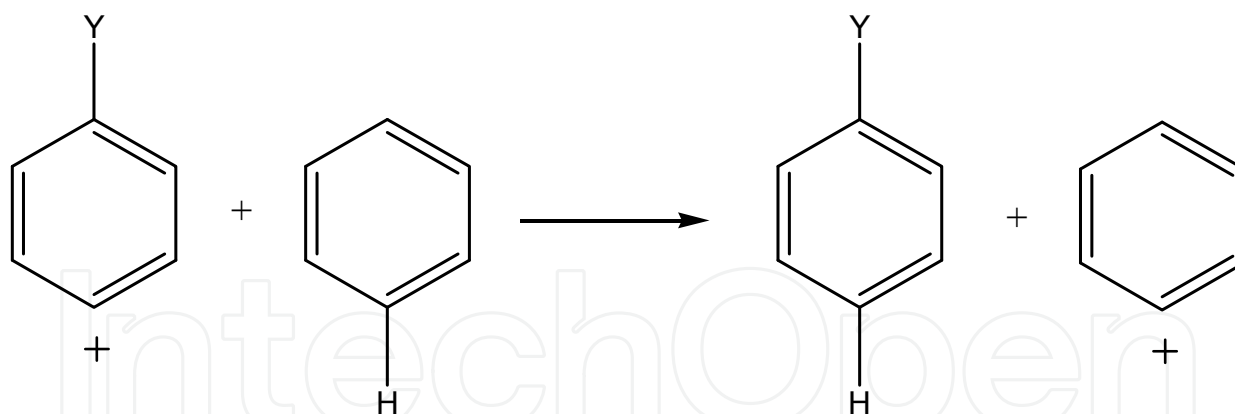
Fig. 9. Isodesmic reaction for the determination of the substituted phenyl cations stability

The calculation of the free energy (ΔG) of this reaction gives a direct information about the stabilization induced by substituent Y. If the free energy is negative for the reaction considered from the left side to the right side, as indicated in the equation the substituted phenyl cation is less stable than the parent cation. If, on the other hand, the reaction has a positive free energy value, the group Y stabilizes the phenyl cation.

If there is a relationship between the geometric classification and the stabilization of the phenyl cation, a pattern should appear by placing in the SOM table the calculated values of free energies. The result is shown in Figure 10, where, for each element of the table, a sphere was drawn having a radius directly proportional to the absolute value of the free energy.

The colour of the sphere indicate the sign of the free energy, blue and red for negative and positive values, respectively. When multiple data are located in the same cell, the sphere is shifted on the right.

As one can see, only a loose correlation between stabilization energy and geometric parameters can be perceived.

4. Conclusion

The geometry and the energy of the singlet and triplet state of a series of substituted phenyl cations have been investigated by DFT methods. The pattern of the effects with reference to the nature of the substituent and to its position has been recognized through the SOM method. The regular hexagon geometry of ring carbons in the triplet state is minimally affected by substituents. The singlet cation is more heavily deformed (cumulene character at the C2-C1-C6 moiety) and is affected to a large extent both by the position and by the nature of the substituent in particular with electron-donating substituents that cause a ring carbon to shift out of plane.
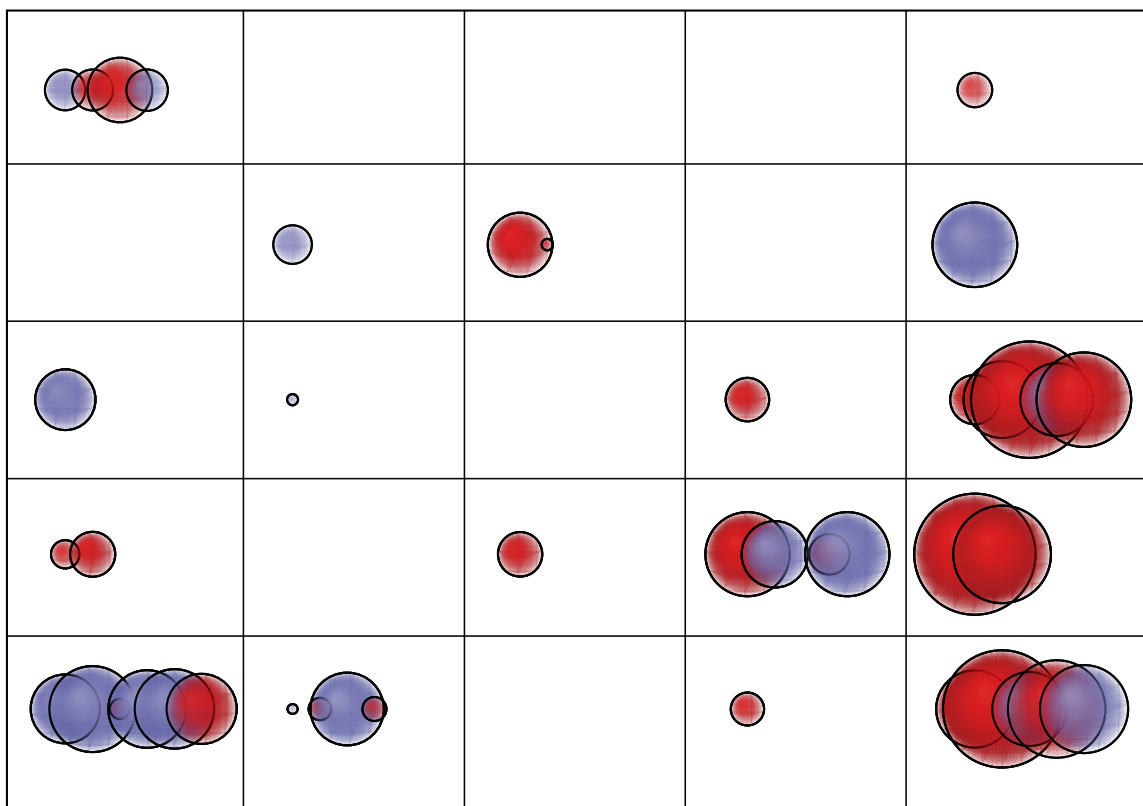

Fig. 10. Stabilization energies of phenyl cations corresponding to the classification in Table 2 (See text)

It was attempted to find a correlation between geometric structure and stabilization energy by the use of this procedure, but this turned out to be weak. We plan to take further advantage of the SOM method for the analysis of the structure and chemical properties of these novel intermediates, phenyl cations.

More generally, we feel that the SOM method should be more largely used for predicting chemical reactivity, both for novel intermediates, as is the case here, and for old ones, where

perhaps it may reveal that the substituent effect is treated in a simplicistic way when using a traditional approach.

5. References

Alonso, M. & Herradin, B. (2007). Neural Networks as a Tool To Classify Compounds According to Aromaticity Criteria, *Chem. Eur. J.*, 13, 3913 – 3923

Bienfait, B. (1994). Applications of High-Resolution Self-Organizing Maps to Retrosynthetic and QSAR Analysis, *J. Chem. Inf. Comp. Sci.*, 34, 890-898

Chen, L. & Gasteiger, J. (1997). Knowledge Discovery in Reaction Databases: Landscaping Organic Reactions by a Self-Organizing Neural Network, *J. Am. Chem. Soc.*, 119, 4033-4042

Dow, L. K.: Kalelkar, S. & Dow, E. R. (2004). Self-organizing maps for the analysis of NMR spectra, *Biosilico*, 2, 157-163

Fagnoni, M. & Albini, A. (2005). Arylation reactions: the photo S 1 path via phenyl cations as an alternative to metal catalysis, *Acc. Chem. Res.*, 38, 713-721.

Fernàndez-Varela, R.; Gòmez-Carracedo, M. P.; Ballabio, D.; Andrade, J. M.; Consonni, V. & Todeschini, R. (2010). Self Organizing Maps for Analysis of Polycyclic Aromatic Hydrocarbons 3-Way Data from Spilled Oils, *Anal. Chem.*, 82, 4264-4271

Gramatica, P. (2007). Principles of QSAR models validation: internal and external Principles of QSAR models validation: internal and external, *QSAR & Comb. Sci.*, 26, 694-701,

Herrero, J.; & Dopazo, J. (2002) Combining Hierarchical Clustering and Self-Organizing Maps for Exploratory Analysis of Gene Expression Patterns, *J. Proteome Res.*, 1, 467-470

Hyvönen, M. T.; Hiltunen, Y.; El-Deredy, W.; Ojala, T.; Vaara, J.; Kovanen, P. Y. & Ala-Korpela, M. (2001). Application of Self-Organizing Maps in Conformational Analysis of Lipids, *J. Am. Chem. Soc.*, 123, 810-816

Kohonen, T.; Hynninen, J.; Kangas, J. & Laaksonen, J. (1996). SOM_PAK: The Self-Organizing Map Program Package. Technical Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science, FIN-02150 Espoo, Finland.

Kohonen T. (2001). *Self Organizing Maps*, Springer-Verlag, Berlin

Lazzaroni, S.; Dondi, D.; Fagnoni, M. & Albini, A. (2008). Geometry and Energy of substituted phenyl cations, *J. Org. Chem.*, 73, 206-211.

Lazzaroni, S.; Dondi, D.; Fagnoni, M. & Albini, A. (2010). Selectivity in the reaction of triplet phenyl cations, *J. Org. Chem.*, 75, 315-323.

Manet, I.; Monti, S.; Grabner, G.; Protti, S.; Dondi, D.; Dichiarante, V. & Albini, A. (2008). Revealing phenylium, phenonium, vinylene phenonium and benzenium ions in solution, *Chem.Eur. J.*, 14. 1029-1039.

Lloyd, G. R.; Brereton, G. R. & Duncan, J. C. (2008). Self Organising Maps for distinguishing polymer groups using thermal response curves obtained by dynamic mechanical analysis, *Analyst*, 133, 1046–1059

Noeske, T.; Sasse, B. C.; Stark, H.; Parsons, C. G.; Weil, T. & Schneider, G. (2006). Predicting Compound Selectivity by Self-Organizing Maps: Cross-Activities ofMetabotropic Glutamate Receptor Antagonists, *ChemMedChem*, 1, 1066 – 1068

Villmann, T.; Schleif, F. M.; Kostrzewa, M.; Walch, A. & Hammer, B. (2008). Classification of mass-spectrometric data in clinical proteomics using learning vector quantization methods, *Brief. Bioinform.*, 9, 129-143

Wehrens, R. & Buydens, L. M. C. (2007) Self- and Super-organizing Maps in R: The kohonen Package, *J. Stat. Soft.*, 21, 1-15

Willighagen, E. L.: Wehrens, R.; Melssen, W.; de Gelder, R. & Buydens, L. M. C. (2007). Supervised Self-Organizing Maps in Crystal Property and Structure Prediction, *Cryst. Gro. Des.*, 7, 1738-1745

Self Organizing Maps - Applications and Novel Algorithm Design

Edited by Dr Josphat Igadwa Mwasiagi

Kohonen Self Organizing Maps (SOM) has found application in practical all fields, especially those which tend to handle high dimensional data. SOM can be used for the clustering of genes in the medical field, the study of multi-media and web based contents and in the transportation industry, just to name a few. Apart from the aforementioned areas this book also covers the study of complex data found in meteorological and remotely sensed images acquired using satellite sensing. Data management and envelopment analysis has also been covered. The application of SOM in mechanical and manufacturing engineering forms another important area of this book. The final section of this book, addresses the design and application of novel variants of SOM algorithms.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Daniele Dondi, Armando Buttafava and Angelo Albini (2011). Application of Self-Organizing Maps in Chemistry. The Case of Phenyl Cations, Self Organizing Maps - Applications and Novel Algorithm Design, Dr Josphat Igadwa Mwasiagi (Ed.), ISBN: 978-953-307-546-4, InTech, Available from:

http://www.intechopen.com/books/self-organizing-maps-applications-and-novel-algorithm-design/application-of-self-organizing-maps-in-chemistry-the-case-of-phenyl-cations

INTECH
open science | open minds