A UNIFIED FORMULA FOR HYDROCARBONS WITH APPLICATIONS TO FUNCTIONAL GROUPS

Higo B. Nogueira*, Norberto de Kássio V. Monteiro b,* and Pedro Lima-Neto b
\textsuperscript{a}Departamento de Física, Universidade Federal do Ceará, 60440-900 Fortaleza – CE, Brasil
\textsuperscript{b}Departamento de Química, Universidade Federal do Ceará, 60020-181 Fortaleza – CE, Brasil

Received em 17/02/2020; aceito em 15/04/2020; publicado na web em 29/05/2020

INTRODUCTION

The Hydrocarbons have a variety of shapes and possible combinations. Even though it is a bielementar molecule, it can present some complexity due to the properties of its constituent elements, in particular to the number of carbon bonds.

This wide range of possibilities has made the scientists look for patterns in the relationship between carbon and hydrogen atoms, even writing some general molecular formulas for certain types of chains. For example, for an straight-chain alkane \( C_xH_y \), we have the relation:  
\[
y = 2x + 2 \quad (1)
\]

There has been also discovered that chains with different properties may contain the same ratio between these two elements. When two or more hydrocarbons have the same general molecular formula, such as alkynes and alkadienes, we can say that they are isomers in this sense. This symmetry of some relationships for different types of chains helps to decrease the total number of general formulas. About this number, we propose the first equation.

Whereas \( H \) is the number of known hydrocarbons types, \( N_i \) is the number of formulas that have the same amount \( i \) of isomers (as defined above), then the number of existing formulas, \( L \), will be given by:

\[
L = H - \sum_i N_i (i - 1) \quad (2)
\]

We can analyze its validity by checking tow limiting cases. If there were no two, or more, isomeric compounds, \( i \) would always be 1 and we would have \( L = H \). At the other limiting case, all hydrocarbons can be considered isomers (as possessing the same formula), thus \( i = H \) and \( N_i = 1 \),

\[
L = H - 1(H - 1) \quad (3)
\]

Therefore,

\[
L = 1 \quad (4)
\]

Here I present the result of (4), which is a single equation that encompasses all molecules composed of carbon and hydrogen, regardless of the shape of their chains.

UNIFIED EQUATION

The construction of this relationship occurred in an empirical way, in which several tests have revealed the usefulness and relevance of specific quantities. First, those quantities will be defined and then the conjecture will be presented.

The type of bond between two carbons will be represented by the letter \( m \), so that: for a single bond, \( m = 1 \); for a double bond, \( m = 2 \), and so on. The equation also presents a summation in which the letter \( m \) appears both as number and index. As an index, \( m \) is attached to \( K_m \) which represents how many times the bonding or order \( m \) appears. For example, the \( K_1 \) for an alkene would be 1; \( K_2 = 1 \) (alkene). This because the double bond (\( m = 2 \)) occurs one time (\( K_2 = 1 \)). Now, we will formalize it.

\[ m = (1,2,3,4) \quad (5) \]

\textbf{Definition 1}: \( m \) is the bonding order between two carbons and can assume four values

\[ K_m \]

\textbf{Definition 2}: \( K_m \) is the number of times the bonding order \( m \) appears in the chain.

\textbf{Definition 3}: \( n \) is the number of cycles in the chain.

\textbf{Conjecture}: Being \( C_xH_y \), any hydrocarbon. The number \( x \) of carbon atoms is related to the number \( y \) of hydrogen atoms by means of the following equation:

\[ x = \frac{y + 2n - 2}{2} + \sum_{m=1}^{4} K_m (m - 1) \quad (6) \]

Before expanding the equation above, we must to assert its limits of applicability. But first, we want to address the new educational tool the equation can be. It may be presented before the main classifications of hydrocarbons, and therefore may help with this. The student who is already familiar with the simple terms in the
equation (such as $K_{\alpha}$ or $x$) is more likely to retain the hydrocarbon’s types and nomenclatures. Furthermore, it would be a more formal way to describe the chain properties.

As a example, the teacher could ask what conditions are necessary for a chain to be aromatic. The most general answer would come in terms of the equations and would be $n \geq 1$, $x \geq 6$, and $K_{\alpha} \geq 3$. As if asking about the conditions for the compound to be of the alkane type: $K_{\alpha}$ other than zero and $K_{\beta} = K_{\gamma} = K_{\delta} = 0$ and so on.

Even though it is complete by itself, there is an optional setting that will be shown in the next section.

**DIMENSIONALITY CONSTANT**

As presented in (6), describes only two-dimensional structures. It means that, if we choose any set of carbon atoms bonded to each other, we will not find closed, volumetric pieces. In the structure form, think of carbon atoms as being vertices and their bindings as being sides of geometric objects. If there is an object, so defined, with non-zero volume, we say that the chain is three-dimensional, $D = 3$. If not, it is a two-dimensional one, $D = 2$.

Taking this into account, it is possible to define the dimensionality constant $C$ as follows

$$C = 2D - 2 \quad (7)$$

The two possible solutions are:

$$C = (2, 4) \quad (8)$$

Equation (2.2) then takes its more general form, adjustable and applicable to all conditions:

$$x = \frac{y + 2n - C}{2} + \sum_{m=1}^{4} K_{\alpha} (m - 1) \quad (9)$$

For the purposes of this work, we used the hydrocarbon equation in its two-dimensional form, $C = 2$, as shown in the previous section.

**INCLUSION OF THE CHALCOCOGENS**

Among the elements that can be added to (6), those of the group 16 are the ones that adapt more naturally. Even so, details about their interactions in the molecule do not appear explicitly in the expression.

Analogously to Section Unified Equation, we will start with some relevant definitions and then the equation will be introduced.

**Definition 4**: $\alpha$ is the bonding order between a chalcogen and a carbon; it can assume the values 1 and 2.

**Definition 5**: $R_{\omega}$ is the number of times that the link of order $\alpha$ appears in the chain.

**Conjecture**: in a molecule $C_{\alpha}H_{\beta}V_{\gamma}$, where $V$ is any element of the group 16, the number of carbons and hydrogens are related according to the following equation:

$$x = \frac{y + 2n - 2}{2} + \sum_{m=1}^{4} K_{\alpha} (m - 1) + \sum_{\omega=1}^{\omega} \frac{R_{\omega}}{2} (\alpha - 1) \quad (10)$$

The similarity between the summation terms is notorious and suggests that the inclusion of other families would occur through the use of other similar terms. Unfortunately, this is not true. However, it is possible to insert the number of atoms $z$, but it happens with some damage to the beauty and simplicity of the expression. This decline in aesthetics happens, among other reasons, by the addition of arbitrary variables. Nevertheless, in the next section we will suggest this relationship.

**THE NUMBER $z$**

The number $z$ does not appear explicitly in (6), but it can be inserted by hand. For this, we need another definition and a conjecture

**Definition 6**: $\omega$ is the number of heteroatoms of element $V$ in the chain.

**Conjecture**: the number $z$ is given by

$$z = \sum_{\omega=1}^{\omega} R_{\omega} + \omega \quad (11)$$

Here it is important to say that the bonds of the heteroatoms $\omega$ do not count as $R_{\alpha}$. By isolating $R_{\alpha}$ in the expression above, it is possible to obtain

$$R_{\alpha} = z - R_{\alpha} - \omega \quad (12)$$

Replacing this value into (10), it is possible to get as result:

$$x = \frac{y + 2n + 2z - 2}{2} + \sum_{m=1}^{4} K_{\alpha} (m - 1) - R_{\alpha} - \omega \quad (13)$$

Equations (10) and (13) are equivalent and produce the same results. Choosing between these two equations is a matter of convenience. We will demonstrate the use of both in section Application to Functional Groups.

**REDUCING THE UNIFIED EQUATION FOR KNOWN GENERAL FORMULAS**

As aforementioned, there are already well-known general formulas for the most common hydrocarbons. Even though there is no formal proof for (6), we can show that it is reduced to such formulas when applied to the respective particular cases. In the following subsections we will do this check for some types of chains and show the reader how to do it in any other cases of interest.

**Alkanes**

Alkanes are defined as open-chain and saturated, branched or normal molecules.\(^{3}\) By conveying this information to the equation, we obtain:

$$x = \frac{y + 2n - 2}{2} + \sum_{m=1}^{4} K_{\alpha} (m - 1) + 0(2 - 1) + 0(3 - 1) \quad (14)$$

Then,

$$y = 2x + 2 \quad (15)$$

which is the general formula for alkanes.

**Alkadienes and alkynes**

Alkadienes and alkynes are general molecular formula isomers, they have the same proportion between Carbon and Hydrogen atoms. The alkadienes are characterized by having two carbon-to-carbon double bonds, while the alkynes have one triple bond. Both are acyclic. Let’s see how the equation behaves in both cases.
For alkadienes:

\[ x = \frac{y + 2.0 - 2}{2} + K'_i (1-1) + 2(2-1) + 0(3-1) \]  

(16)

and for alkynes:

\[ x = \frac{y + 2.0 - 2}{2} + K'_i (1-1) + 0(2-1) + 1(3-1) \]  

(17)

The two equations above generate the same formula, notably the general molecular formula for alkadienes and alkynes.

\[ y = 2x - 2 \]  

(18)

**Cycloalkanes**

Cycloalkanes are compounds that have one cycle and a certain number of single bonds. Thus, we find its general formula:

\[ x = \frac{y + 2.1 - 2}{2} + K'_i (1-1) + 0(2-1) + 0(3-1) \]  

(19)

thus,

\[ y = 2x - 2 \]  

(20)

**APPLICATION TO FUNCTIONAL GROUPS**

Now it is possible to see which form the relations assume when one more element is introduced. We will do it for some functional groups containing oxygen, using (10) and (13).

**Ether and Phenol**

These two functional groups generate the same equation, surprisingly the hydrocarbon equation. First, we will do the assembly for the Ether and then for the Phenol. In both processes the Eq. (13) will be used:

\[ x = \frac{y + 2(1) + 2n - 2}{2} + \sum_{m=1}^{4} K'_m (m-1) - 0 - 1 \]  

(21)

also,

\[ x = \frac{y + 2(1) + 2n - 2}{2} + \sum_{m=1}^{4} K'_m (m-1) - 1 - 0 \]  

(22)

The two expressions above are reduced to the known Equation (3.3)

\[ x = \frac{y + 2n - 2}{2} + \sum_{m=1}^{4} K'_m (m-1) \]  

(23)

**Ketone and Aldehyde**

Ketone and aldehyde also generate the same final equation, in this particular case even the intermediate equations are the same. The following assembly is used for the two functions

\[ x = \frac{y + 2n - 2}{2} + \sum_{m=1}^{4} K'_m (m-1) + 1 \]  

(24)

and again, we get one single result:

\[ x = \frac{y + 2n}{2} + \sum_{m=1}^{4} K'_m (m-1) \]  

(25)

**GENERALIZATION AND PERSPECTIVES**

With a unified formula for hydrocarbons at hand, we can postulate even bigger ones. Thus, we will have two fronts of generalization: the inclusion of a third element (in section 4 we included chalcogens, albeit in a rudimentary way); and the exchange of hydrogen and oxygen for any two elements, making the equation valid for any molecule composed of two elements.

In the first case, we have a pragmatic interest in studying algebraically the functional groups that would be described by the generalized equation. The second case represents a more formal, abstract, but not less important interest. Understanding how the equation would work and what form it would take for any two elements constitutes a fundamental step to postulate an equation for any molecule composed of three elements, which in turn encompasses cases of greater interest and applicability.

Following this line of reasoning to its end, we can conjecture the existence of an equation that describes any molecule, regardless of the quantity or types of elements.

An important and parallel step to the others is the parameterization of the formulas by the bonding energies, causing their solutions to be dictated by the expected and initial values of the system’s energy. Other parameterizations may also be possible and useful for different cases.

**CONCLUSIONS**

Equation (6) is a formula containing many other formulas by itself. It synthesizes a variety of particular, previously dispersed relationships. It can also be programmed and can provide any kind of information of interest to the researcher; and it is simple enough to be learned and used by high school students. From it, we can generate more powerful equations through the inclusion of entire families of elements. The equation itself is a non-trivial piece of truth; the interested researcher might add particularities of his study field to the equation, in order to make it applicable.

It also represents an effort of language, by turning words into numbers. This effort must be followed and broadened as indicated by the three path suggestions pointed out in the previous section. The intention of this study is to establish and provide ways to produce equations such that anyone can solve problems by them. In a way that, to understand the state and the evolution of a system, it will be enough to solve one equation.

**ACKNOWLEDGMENT**

Authors Thank Coordenação de Ensino de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for financial support.

**REFERENCES**

1. Speight, J. G.; *Handbook of Industrial Hydrocarbon Processes*, Elsevier: Amsterdam, 2011.
2. Mackay, D.; Shiu, Y. W.; Ma, C. K.; Lee, C. S.; *Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals*, 2nd ed., Broken Sound Parkway: New York, 2006.
3. Carey, A. F.; Sundberg, J. R.; *Advanced Organic Chemistry: Part A: Structure and Mechanisms*, 5th ed., Spring Street: New York, 2007.