



Rényi's divergence as a chemical similarity criterion

N. Flores-Gallegos¹

Received: 9 August 2021 / Accepted: 9 November 2021 / Published online: 22 November 2021

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

Abstract

In this work, a new version of Rényi's divergence is presented. The expression obtained is used as a tool to identify molecules that could share some chemical or structural properties, and a data basis set of 1641 molecules is used in this study. Our results suggest that this new form of Rényi divergence could be a useful tool that will eventually permit complementary studies in which the main goal is to obtain molecules with similar properties.

Keywords Rényi's entropy · Rényi's divergence · Chemical similarity · Chemical reactivity

1 Introduction

Perhaps, currently in the modern sciences, there are no clear barriers between fields; for example, let us consider two different areas such as biology and computer sciences which in the last decade formed a field called bioinformatics. A similar situation happens with chemistry, in where there are no clear separations between the modern chemistry, physics, mathematics, and computer sciences; in this regard, a new discipline of chemistry called cheminformatics or molecular informatics has emerged in recent years. In this area, one of the objectives is to find molecules that may have some chemical or physical properties of interest.

In the last year, in which, for practically the entire population of the world, it was suggested to stay isolated at home, all the scientific community was called to participate in the '*JEDI Grand Challenge: Billion molecules against COVID-19*'. In an announcement by The Joint European Disruptive Initiative that was published on the main page of the journal '*Molecular Informatics*', one can read, '*The Joint European Disruptive Initiative has announced a May 1st launch of the Billion Molecules against Covid19 Grand Challenge, with awards of up to € 2 million for the winners.*'

✉ N. Flores-Gallegos
nelson.flores@academicos.udg.mx

¹ Centro Universitario de los Valles, Universidad de Guadalajara, Carretera Guadalajara - Ameca Km. 45.5, C.P. 46600 Ameca, Jalisco, Mexico

for which the main objective is ‘*to screen billions of molecules with blocking interactions relevant to SARS-CoV-2 and fast-track the route to a therapeutic treatment.*’

It is interesting to note that to address a problem such as the mentioned before, there are two necessary ingredients; the first, of course, is the equations, which must be defined using solid fundamentals, instead of empirical, heuristic, or statistical assumptions, and the second is the algorithms, that is, how can we improve the search of biological, chemical, or physical properties that a group of molecules could have in common to obtain a biological or chemical agent that permits to define of a set of molecules that may act as a good treatment or even a vaccine.

In this sense, cheminformatics is an area that has a strong link with QSAR analysis [1]; however, it is not common to find that such analyses are complemented by descriptors based on the formal fundamentals of theoretical chemistry. In this regard, perhaps the first proposal of a similarity coefficient based on electron density was done in the late 1970s by Carbó [2].

On the other hand, according to Willett [3], a cheminformatics analysis has three components, (i) a structure representation, (ii) a weighting scheme, and (iii) a similarity coefficient; in the first case, the analysis consists of performing a comparison of the structure between two molecules, using as a criterion the structural representation of the molecules in 2D or 3D; in the second case, one has to define a fingerprint of several molecules considering the number of the apparition of certain functionals groups, and in the last case, the analysis consists of choosing certain coefficients defined in terms of the number of features present in the compounds to analyze. In this sense, clearly one can define a vast number of such coefficients [4]; for example, in Todeschini’s work, the authors analyzed 51 different coefficients [5], which could be an attractive aspect to perform some studies in the field of medicinal chemistry [6]. However, it is worthwhile to mention that the majority of these coefficients were defined in a heuristic way.

In this work, we propose the use of a new definition of Rényi’s divergence that can be used as a tool to find molecules that could have some chemical or structural properties in common.

In the following section, we present a brief background of Rényi’s divergence and its mathematical properties.

2 Theoretical background

Since the publication of Shannon’s work in the 1940s, information theory [7] has been applied in several areas of science and technology. In chemistry, the use of the ideas of information theory started in the 1970s, and these studies were principally focused on the study of atoms and simple systems using Shannon’s entropy in position and momentum spaces; in the 1990s, with the advantages of computers, it was possible to extend these studies to simple molecules and some simple chemical processes. In this line, it was also possible to show that Shannon’s entropy can be related to some important concepts of quantum chemistry, such as the chemical reactivity, electron correlation, and structural stability of the molecules [8–17].

On the other hand, in the 1960s, the idea of generalizing Shannon's entropy emerged, and today, one can find three generalizations of it: (i) Rényi's entropy [18] (ii) the Havrda and Charvát entropy [19] and (iii) Tsallis' entropy [20]. The first two were proposed in the 1960s, while the third one was proposed at the end of the 1980s.

Considering a historical point of view, perhaps Rényi's work can be considered the first attempt to generalize Shannon's entropy. Since the 1990s, Rényi's entropy has been profusely used in the field of atomic and molecular physics [21–36], and these works were focused on the study of the electron correlation phenomenon.

Nevertheless, we would like to address the attention of the readers to another concept, also proposed by Rényi, which is the divergence, and it can be interpreted as a measure that compares and quantifies the difference between two probabilistic distributions. It is defined as

$$R_{ij,\alpha} = \frac{1}{\alpha - 1} \ln \left(\frac{1}{k_{nm}} \sum_{i=1}^n \sum_{j=1}^m \frac{p_i^\alpha}{q_j^{\alpha-1}} \right), \quad (1)$$

and

$$R_{ji,\alpha} = \frac{1}{\alpha - 1} \ln \left(\frac{1}{k_{nm}} \sum_{i=1}^n \sum_{j=1}^m \frac{q_j^\alpha}{p_i^{\alpha-1}} \right), \quad (2)$$

in which the sets $\{p_i\}$ and $\{q_j\}$ are probabilistic distributions that fulfill $0 \leq p_i \leq 1$, $0 \leq q_j \leq 1$, and $\sum_{i=1}^n p_i = 1$, $\sum_{j=1}^m q_j = 1$. In both cases, the term α , can take values of $\alpha = 0, 1, \dots$, but with the exception that $\alpha \neq 1$. Here, $k_{nm} = nm$, and n and m are the sizes of the sets $\{p_i\}$ and $\{q_j\}$, respectively. These constants prevent the numerical overestimation of the sums (see “Appendix”).

From Eqs. (1) and (2), one can infer easily that $R_{ij,\alpha} \neq R_{ji,\alpha}$; because of this, we need to ensure that the expression used to calculate the divergence is symmetric; therefore, our proposal for the Rényi's divergence is

$$R_\alpha = R_{ij,\alpha} + R_{ji,\alpha}, \\ = \frac{1}{\alpha - 1} \left(\ln \frac{1}{k_{nm}} \sum_{i=1}^n \sum_{j=1}^m p_i^\alpha q_j^{1-\alpha} + \ln \frac{1}{k_{nm}} \sum_{i=1}^n \sum_{j=1}^m p_i^{1-\alpha} q_j^\alpha \right). \quad (3)$$

If one applies the logarithm properties to this equation, then it can be reduced to

$$R_\alpha = \frac{1}{\alpha - 1} \ln \left(\sum_{i=1}^n \sum_{j=1}^m \frac{p_i q_j}{k_{nm}} \right), \quad (4)$$

where $k_{nm} = nm$; Eq. (3) and consequently Eq. (4) both fulfill the following criteria [37]:

1. R_α is a continuous function of p_1, p_2, \dots, p_n and q_1, q_2, \dots, q_m .

2. R_α is permutationally symmetric; that is, the total value of this measure does not change if the elements p_1, p_2, \dots, p_n and q_1, q_2, \dots, q_m are labeled in a different way, which implies that the pairs $(p_1, q_1), (p_2, q_2), \dots$ can permute among themselves.
3. $R_\alpha \geq 0$.
4. Based on the previous property, then one can infer that the minimum value of R_α is zero.

In the field of information theory, it is well known that Rényi's entropy is a generalization of Shannon's entropy; in this regard, Rényi's divergence is a generalization of several divergence measures. For example, if one considers $\alpha = 0$, Eq. (4) becomes,

$$R_{\alpha=0} = -\ln \left(\sum_{i=1}^n \sum_{j=1}^m \frac{p_i q_j}{k_{nm}} \right). \quad (5)$$

An interesting aspect of this equation is that if one applies the exponential function to obtain $e^{-R_{\alpha=0}}$, one obtains a discrete version of Carbo's similarity coefficient, which is related to the overlap of two distributions [38–41].

If we consider $\alpha = 1/2$, Eq. (3) becomes,

$$R_{\alpha=1/2} = -4 \ln \sum_{i=1}^n \sum_{j=1}^m \frac{\sqrt{p_i q_j}}{k_{nm}}, \quad (6)$$

which is known as the Bhattacharyya coefficient [42] and has been used extensively in the field of computational sciences as a measure of overlap between two statistical distributions.

When $\alpha \rightarrow 1$, Eq. (3) converges to the Kullback–Leibler entropy [43],

$$R_{\alpha \rightarrow 1} = \frac{1}{k_{nm}} \sum_{i=1}^n \sum_{j=1}^m p_i \ln \frac{p_i}{q_j} + \frac{1}{k_{nm}} \sum_{i=1}^n \sum_{j=1}^m p_j \ln \frac{p_i}{q_j}. \quad (7)$$

Other interesting aspect of Eq. (3) appears when $\alpha = 2$; in this case, Eq. (3) has the following form,

$$\begin{aligned} R_{\alpha=2} &= \ln \frac{1}{k_{nm}} \sum_{i=1}^n \sum_{j=1}^m \frac{p_i^2}{q_j} + \ln \frac{1}{k_{nm}} \sum_{i=1}^n \sum_{j=1}^m \frac{q_j^2}{p_i}, \\ &= \ln \left(\sum_{i=1}^n \sum_{j=1}^m \frac{p_i q_j}{k_{nm}} \right), \end{aligned} \quad (8)$$

in this equation, one can recognize the terms $\frac{p_i^2}{q_j}$ and $\frac{q_j^2}{p_i}$ as terms associated with the chi-squared distribution, which is a useful tool in statistical analyses.

In general, we can say that Eqs. (5), (6), (7) and (8) can be used as measures that permit the comparison of two different probabilistic distributions.

In this regard, if one extrapolates these equations to the context of quantum chemistry, then one obtains a set of so-called quantum similarity coefficients. Thus, if one wants to express Eq. (4) in its quantum version, then it is necessary to rewrite it in terms of two diagonalized density matrices, namely, ρ_A and ρ_B . These density matrices are related to the quantum states of the systems A and B ; therefore, we can write Eq. (4) in terms of the trace of ρ_A and ρ_B to obtain

$${}^qR_\alpha = \frac{1}{\alpha - 1} \left[\ln \frac{1}{k_{nm}} \text{Tr}_A \rho_A^\alpha \text{Tr}_B \rho_B^{1-\alpha} + \ln \frac{1}{k_{nm}} \text{Tr}_A \rho_A^{1-\alpha} \text{Tr}_B \rho_B^\alpha \right], \quad (9)$$

where the superscript indicates the quantum version of Eq. (4), which can be rewritten in the following compact and simplest form:

$${}^qR_\alpha = \frac{1}{\alpha - 1} \ln \left[\frac{1}{k_{nm}} (\text{Tr}_A \rho_A) (\text{Tr}_B \rho_B) \right]. \quad (10)$$

To apply Eq. (9) to systems such as atoms, molecules, macromolecules, it is necessary to use two diagonalized density matrices, in which each element of their trace is an occupation number; in the theoretical chemistry field, the occupation numbers are subject to $0 \leq \lambda_k \leq 2$, $\sum_{k=1}^I \lambda_k = N$, where N is the electron number. It is convenient to clarify that such numbers are restricted to the values of 0, 1, 2 when they are obtained through mono-determinantal methodologies, such as HF or DFT; moreover, when they are obtained through multideterminantal methodologies as CI or CC, they are subject to $0 \leq \lambda_k \leq 2$, where $\lambda_k \in \mathbb{R}$ and $\sum_{k=1}^I \lambda_k = N$. Therefore, to use such numbers with Eq. (10) they have to be normalized as $\lambda'_k = \lambda_k/N$. This new set of numbers is subject to $0 \leq \lambda'_k \leq 1$, and $\sum_{k=1}^I \lambda'_k = 1$, and now the set $\{\lambda'_k\}$ fulfills all criteria of a probabilistic distribution, which is one of the main ingredients of any information measure.

Perhaps the most interesting aspect of Eq. (10) is that it permits comparison of the quantum information of two molecular systems, which are described by two density matrices, and such matrices, according to Dirac, completely describe the state of the quantum systems [44]. In this sense, it is not hard to see that Eq. (10) can represent a quantum superposition; that is, in the context of quantum mechanics, a quantum superposition is considered a sum of states to obtain a new state. For example, if one considers Eq. (9) with $\alpha = 0$, it is evident that this equation is reduced to a sum of logarithms, and each logarithm has an argument for the states, ρ_A and ρ_B ; therefore, one can interpret the numerical results obtained with Eq. (10) also as a quantum superposition.

On the other hand, in the context of density functional theory, there are three main concepts linked to chemical reactivity, hardness, softness, and chemical potential [45], which are defined as

$$\text{Hardness} = \eta = \frac{1}{2} \left(\frac{\partial \mu}{\partial N} \right)_v = \frac{1}{2} \left(\frac{\partial^2 E}{\partial N^2} \right)_v, \quad (11)$$

$$\text{Softness} = S = \frac{1}{2\eta} = \left(\frac{\partial N}{\partial \mu} \right)_v, \quad (12)$$

where the chemical potential, μ , is define as

$$\mu = \left(\frac{\partial E}{\partial N} \right)_v = -\chi, \quad (13)$$

and χ is Mulliken's electronegativity definition [46].

Equations (11), (12) and (13) can be rewritten in terms of the electron affinity, A , and the ionization potential, I , as

$$\eta \approx \frac{I - A}{2}, \quad (14)$$

$$S = \frac{2}{I - A} \quad (15)$$

and

$$\mu \approx -\frac{I + A}{2}. \quad (16)$$

Using Koopman's theorem and under the LCAO approximation, the last three equations can be rewritten in terms of the HOMO and LUMO energies as

$$\eta \approx \frac{E_{\text{LUMO}} - E_{\text{HOMO}}}{2} \quad (17)$$

$$S \approx \frac{2}{E_{\text{HOMO}} - E_{\text{LUMO}}}, \quad (18)$$

and

$$\mu \approx -\frac{E_{\text{LUMO}} + E_{\text{HOMO}}}{2}. \quad (19)$$

In general, Eqs. (17), (18) and (19) are used as powerful tools to study how a molecule can react under different chemical conditions; for this reason, we consider that such equations are important to complement our study.

In the following section, we present some numerical results when Eq. (10) is applied to a large set of molecules; our analysis is complemented by the numerical results of Eqs. (17), (18) and (19).

3 Results

To apply and to show the versatility of Eq. (10), a set of 1641 molecules defined by Blair and Thakkar [47] was chosen. Such molecules are closed-shell molecules, and according to the authors, the molecules contain up to 34 atoms and 246 electrons, which can be used to study the interrelationships among the properties of molecules of organic, biochemical, and pharmaceutical interest.

To perform the calculations of this study, we used Eq. (10) with the Löwdin occupation numbers obtained with the CISD method and DGDZVP basis set, while to calculate the hardness, softness and chemical potential in terms of the HOMO and LUMO energies, we used the functional M062x with the DGDZVP basis set. All calculations were performed with Gaussian 09 [48].

For practical proposes, we chose in a random way one molecule of the set of molecules used. The molecule chosen was number 750, tetranitromethane, $C(NO_2)_4$, which is a highly explosive chemical used as an oxidizer in rocket propellants and is used to increase the cetane number of diesel fuels in the manufacture of liquid explosives. It has been classified as possibly carcinogenic to humans [49]. This molecule has 98 electrons, a softness value of $\eta = 0.16588$ [a.u.], a hardness value of $S = 3.0142271$ [a.u.] and a chemical potential value of $\mu = -0.28717$ [a.u.].

In Tables 1, 2, 3 and 4, we present the numerical results of Eq. (10) using the values of $\alpha = 0, 0.5, 1, 2$, and in Figs. 2, 3, 4 and 5 are depicted the molecules that according to the numerical results of Eq. (10) may share some properties. At this point, we consider that it is opportune to mention that we report only the numerical results of the molecules that are similar, instead of the 1641 numerical values of ${}^qR_\alpha$ for each α value considered in this study; however, we present the general trends of ${}^qR_\alpha$ in Fig. 1a–d. For each case, a percentage of difference of ${}^qR_\alpha$ among the molecules was chosen because, for example, if one observes Fig. 1a, in which the location of the molecule 750 is highlighted with red lines and follows the horizontal line, one can appreciate that there is a dense section in which several molecules appear. For that reason, we selected the molecules with a difference of 0.0001%, while for ${}^qR_\alpha$ with $\alpha = 0.5, 1, 2$, the difference used was 0.2%, 0.3% and 0.1%. To obtain small sets of molecules, clearly, one can change these percentage values to increase the size of the sets; on the other hand, if one follows the vertical line on Fig. 1a, one

Table 1 Numerical results of Eq. (10) with $\alpha = 0$

Molecule	N	${}^qR_{\alpha=0}$	η [a.u.]	S [a.u.]	μ [a.u.]
mol750	80	10.4080133741533	0.2073000	2.4119633	-0.098680
mol1041	84	10.4080133741534	0.1286900	3.8853057	-0.151720
mol1163	90	10.4080133741533	0.1561650	3.2017417	-0.150825
mol1553	82	10.4080133741534	0.1472700	3.3951246	-0.126900

The molecule of reference was tetranitromethane with ${}^qR_\alpha = 10.4080133741532$, $N = 98$, $\eta = 0.1658800$, $S = 3.0142271$, and $\mu = -0.287170$. mol750: tetraspiro[2.0.0.0.2.1.1.1]undecane, mol1041: 1,2-furylpyridin-2-one, mol1153: 1H-tetrazol-5-amine, mol1163: 1R-1,2,2-trifluoroethoxybenzene

Table 2 Numerical results of Eq. (10) with $\alpha = 0.5$

Molecule	N	${}^qR_{\alpha=0.5}$	η [a.u.]	S [a.u.]	μ [a.u.]
mol58	62	25.3760755737034	0.0964350	5.1848395	-0.148575
mol86	70	25.368376634063	0.2169000	2.3052097	-0.102550
mol149	70	25.3316736530311	0.1228650	4.0695071	-0.177535
mol153	70	25.369118109399	0.2252700	2.2195587	-0.092870
mol239	76	25.3430902879608	0.1281750	3.9009167	-0.122995
mol567	68	25.3766256442875	0.1268800	3.9407314	-0.143010
mol606	76	25.3405865665757	0.1347500	3.7105751	-0.227120
mol776	70	25.3696698654121	0.2321500	2.1537798	-0.107770
mol815	88	25.3704974979544	0.1518400	3.2929399	-0.183810
mol838	70	25.3310663275822	0.1518550	3.2926146	-0.109345
mol860	74	25.3757483615868	0.2250150	2.2220740	-0.102315
mol948	74	25.3750316177502	0.2254400	2.2178850	-0.116470
mol949	72	25.3791298985497	0.1413600	3.5370684	-0.167100
mol999	82	25.350679562435	0.2232550	2.2395915	-0.112635
mol1132	70	25.3696564444162	0.2247200	2.2249911	-0.112970
mol1335	70	25.3305640125937	0.1327850	3.7654855	-0.100355
mol1487	70	25.3773791334903	0.1342000	3.7257824	-0.117240

The molecule of reference was tetranitromethane with ${}^qR_{\alpha} = 25.3553653711884$, $N = 98$, $\eta = 0.1658800$, $S = 3.0142271$, and $\mu = -0.287170$. Where, mol58: azulene, mol86: bicyclo-3,3,1-nonane, mol149: cinnamaldehyde, mol153: cis-bicyclo-6,1,0-nonane, mol239: difuro[3,2-b:2',3'-d]furan, mol567: naphthalene, mol607: p-dichlorobenzene, mol776: trans-2,3-dimethylnorborane, mol815: vinylsulfonylbenzene, mol838: 1'H-1,2'-bipyrrole, mol860: 1-butoxybutane, mol948: 1-octanol, mol949: 1-p-tolyethanone, mol999: 1,1,1,2-tetramethoxyethane, mol1132: 1,4-dimethylnorborane, mol1335: 2,1H-pyrrol-2-yl-1H-pyrrole, mol1487: 3-methyl-1H-indole

Table 3 Numerical results of Eq. (10) with $\alpha = 1$

Molecule	N	${}^qR_{\alpha=1}$	η [a.u.]	S [a.u.]	μ [a.u.]
mol204	90	10.541193464482	0.1690700	2.9573549	-0.106900
mol563	76	10.8810027096117	0.1176650	4.2493519	-0.126665
mol565	76	10.8832930882276	0.1193050	4.1909391	-0.127765
mol815	88	10.5894488246596	0.1518400	3.2929399	-0.183810
mol879	84	10.6721705605749	0.2106800	2.3732675	-0.142970

The molecule of reference was tetranitromethane with ${}^qR_{\alpha} = 10.378971157751$, $N = 98$ and $\eta = 0.1658800$, $S = 3.0142271$, and $\mu = -0.287170$. mol204: di-t-butyl sulfoxide, mol563: naphthalen-1-amine, mol565: naphthalen-2-amine, mol815: vinylsulfonylbenzene, mol879: 1-chloro-1,3-butadiene

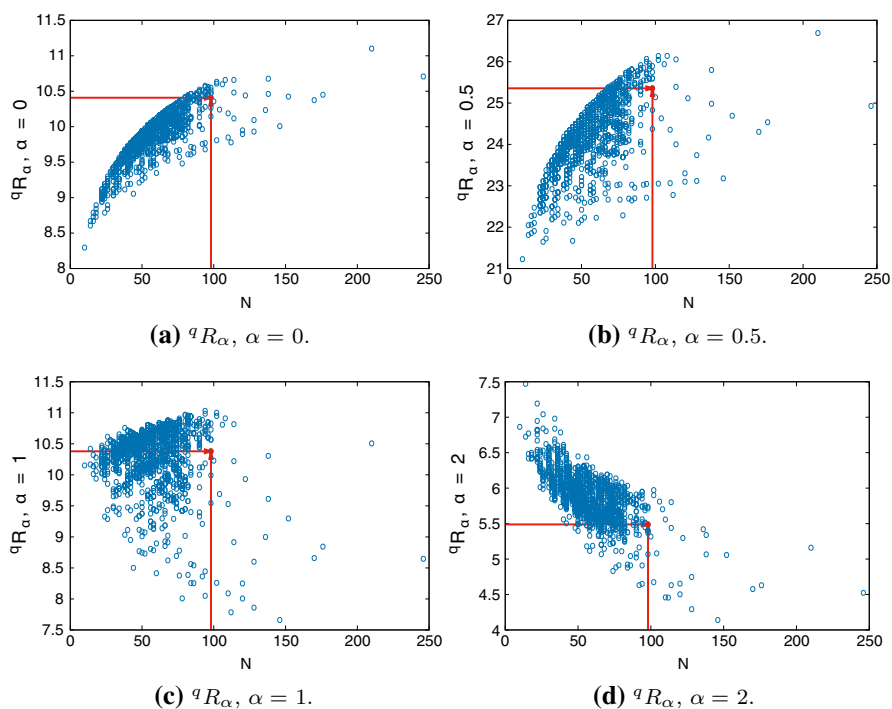
observes that there are several isoelectronic molecules, with $N = 98$, but in the majority of the cases, the value of ${}^qR_{\alpha}$ differentiates all these molecules.

In addition, in Fig. 1a–d, it is also possible to observe the effect of the coefficient α on ${}^qR_{\alpha}$; in such figures, we noticed that for values of $\alpha \in [0, 1]$, the general trends have a convex behavior, while for the interval $\alpha \in (1, 2]$, the trends are concave.

Table 4 Numerical results of Eq. (10) with $\alpha = 2$

Molecule	N	${}^qR_{\alpha=2}$	η (a.u.)	S (a.u.)	μ (a.u.)
mol242	98	5.36699258960823	0.2105750	2.3744509	-0.132485
mol546	90	5.77672947791593	0.1415950	3.5311981	-0.176035
mol879	84	5.91519112233904	0.1238100	4.0384460	-0.151740
mol1640	86	5.81025974210546	0.1360050	3.6763354	-0.146185

The molecule of reference was tetranitromethane with ${}^qR_{\alpha} = 5.48769619636878$, $N = 98$ and $\eta = 0.1658800$, $S = 3.0142271$, and $\mu = -0.287170$. mol242: diisobutyl sulfone, mol546: N-pyrazin-2-ylmethanesulfonamide, mol879: 1-chloronaphthalene, mol1640: 9H-purine-2,6,8-trione

**Fig. 1** Trends of ${}^qR_{\alpha}$ with $\alpha = 0, 0.5, 1, 2$

Therefore we can say that depending on the α value used, we can find more or fewer molecules because some could be or could not be in the same position. This can be inferred easily by comparing the position of the molecule of reference used, which is depicted in the figures mentioned previously.

Observing the four tables, one can note that depending on the α value used, different molecules are found that according to Eq. (10) are informationally equivalent. For example, with $\alpha = 0$ (see Table 1 and Fig. 2), the molecules 750, 1041, 1153 and 1163 are equivalent under the perspective of ${}^qR_{\alpha}$. These molecules have structures constituted mainly by rings, and three of them present aromaticity, while the

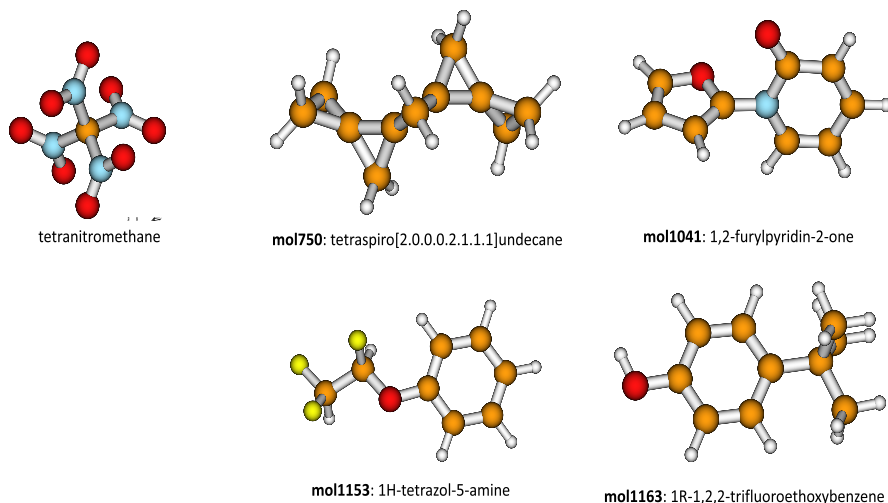


Fig. 2 Graphic representation of the molecules: mol750, mol563, mol565, mol815, mol879, used in Table 1

structure of tetranitromethane is tetrahedral (see Fig. 2). From a chemical reactivity point of view, the molecules 1041, 1153, and 1163 have similar values of softness, hardness, and chemical potential; however, if we compare such values with the values of tetranitromethane, we find that molecule 1163 has the closest value of hardness and softness but that its value of chemical potential differs significantly. Thus, based on the numerical results, we conclude that 1R-1,2,2-trifluoroethoxybenzene is similar to tetranitromethane.

In Table 2 are reported the numerical results of Eq. (10) with $\alpha = 0.5$, and the graphic representation of the molecules of this case is shown in Fig. 3. The first general observation that we note is that with this value of α , it is possible to find more molecules in comparison with the previous case, but in this case, the numerical values of ${}^qR_\alpha$ present a percentage difference of 0.2%. As in the previous case, this set of molecules constitutes mostly rings, but in this case, there are two linear molecules present in the set (see Fig. 3). Now, if we compare the values of softness, hardness, and chemical potential, it is possible to reduce this set of molecules. In this case, we find that only molecules 815 and 838 have similar hardness values $\eta = 0.15184$ [a.u.] and $\eta = 0.151855$ [a.u.] in comparison with the value for tetranitromethane of $\eta = 0.16588$ [a.u.]. A similar situation happens with the hardness; however, the chemical potential of molecules 815 and 838 differs by almost 0.1 a.u. with respect to that of tetranitromethane.

In Tables 3 and 4, we present the results of ${}^qR_\alpha$. Using the values of $\alpha = 1$ and 2, as in the first case, these values of α produce values of ${}^qR_\alpha$ with a small percentage difference. In both sets are present molecules with rings, and all they have some of the most electronegative atoms; nevertheless, in the case of the set with $\alpha = 2$ we also find two linear molecules (see Figs. 4 and 5). When a value of $\alpha = 1$ is considered, only molecules 204 and 815 have similar hardness and softness values, while

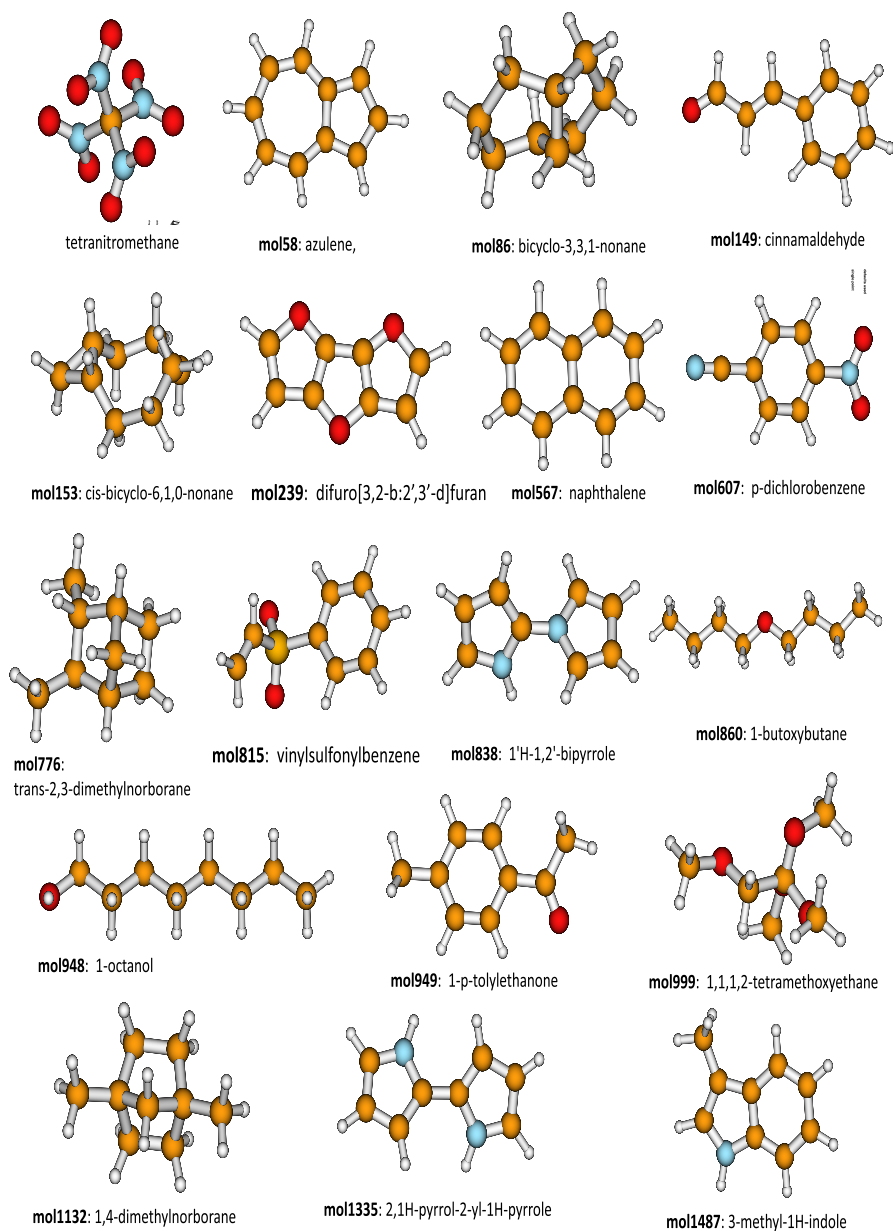


Fig. 3 Graphic representation of the molecules: mol58, mol86, mol149, mol153, mol239, mol567, mol607, mol776, mol815, mol838, mol860, mol948, mol949, mol999, mol1132, mol1335, mol1487, used in Table 2

the chemical potential is different by more than 50% with respect to tetranitromethane. On the other hand, with the value of $\alpha = 2$, the four molecules differ considerably in all their chemical reactivity parameters with respect to tetranitromethane.

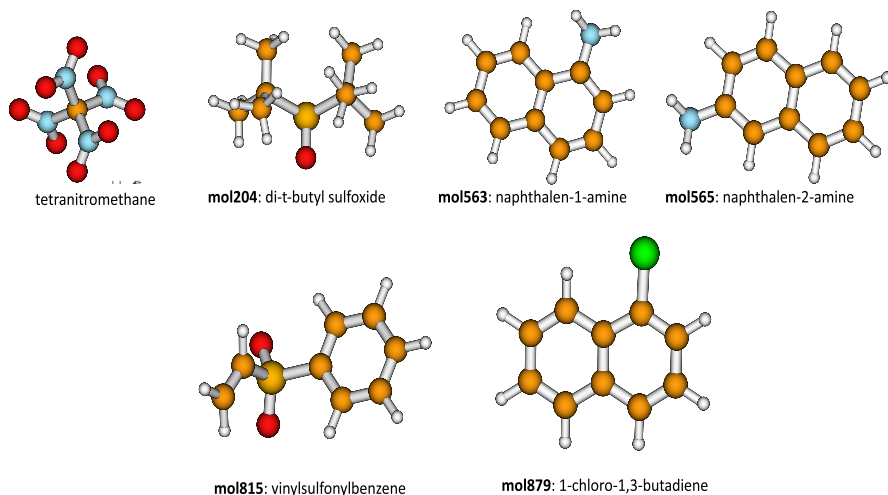


Fig. 4 Graphic representation of the molecules: mol204, mol1041, mol1153, mol1163, used in Table 3

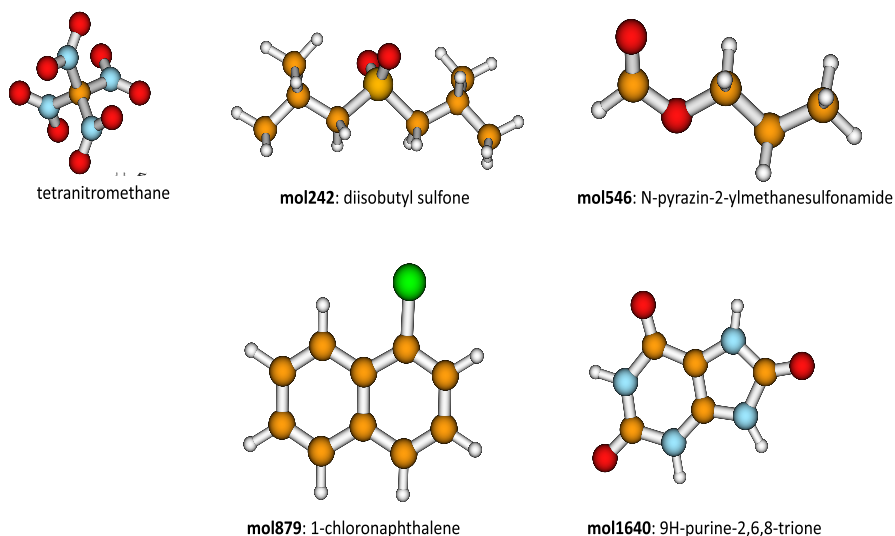


Fig. 5 Graphic representation of the molecules: mol242, mol546, mol879, mol1640, used in Table 4

In Table 5, we summarize our findings of ${}^qR_\alpha$. In this case, it was possible to reduce the set of molecules to four molecules. Based on this table, we note that molecule 204, di-t-butyl sulfoxide, with $\alpha = 1$, has hardness and softness values close to tetranitromethane, which is the molecule of reference. This result may be explained in terms of the structure of both molecules (see Fig. 6). In this figure, one can observe that both molecules are semispherical, while the rest of them have rings and are semiplanar. Another interesting observation of the results presented in Table 5 is that the numerical

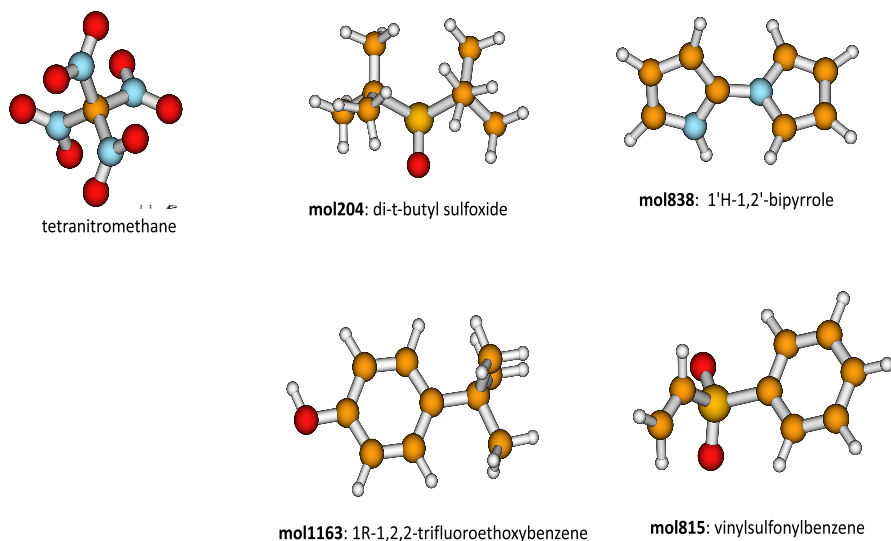


Fig. 6 Graphic representation of the molecules: mol204 mol815, mol838, mol1163, used in Table 5

Table 5 Selection molecules that could share some chemical properties

Molecule	N	${}^qR_\alpha$	α	η [a.u.]	S [a.u.]	μ [a.u.]
Tetranitromethane	98	–	–	0.165880	3.0142271	–0.28717
mol204	90	10.541193464482	1	0.169070	2.9573549	–0.106900
mol815	88	10.5894488246596	1	0.151840	3.2929399	–0.183810
mol815	88	25.3704974979544	0.5	0.151840	3.2929399	–0.183810
mol838	70	25.3310663275822	0.5	0.151855	3.2926146	–0.109345
mol1163	90	10.4080133741533	0	0.156165	3.2017417	–0.150825

mol204: di-t-butyl sulfoxide, mol815: vinylsulfonylbenzene, mol838: 1'H-1,2'-bipyrrole, mol1163: 1R-1,2,2-trifluoroethoxybenzene

results of ${}^qR_\alpha$ with $\alpha = 0.5, 1$ suggest that molecule 815, vinylsulfonylbenzene, has similar properties to tetranitromethane; however, we also note that while the hardness and softness in several cases match with the molecule of reference, the chemical potential does not.

In conclusion, we can say that the version of Rényi's divergence presented in this work could be an effective tool used in conjunction with the descriptors of the chemical reactivity to complement studies where the main propose is to find some chemical systems that may share some properties.

4 Conclusion

In this work, we presented a new version of Rényi's divergence defined in terms of two diagonalized density matrices. The equation proposed in this work satisfies rigorously all the properties of any informational measure. Our results suggest that Eq. (10) with values of $\alpha = 0.5$ and $\alpha = 1$ produces the best sets of molecules that may share some properties.

Finally, we consider that our expression of Rényi's divergence could be a good candidate to complement studies in which the main goal is the search for chemical systems with similar properties or as a tool to classify sets of molecules that may share some physical or chemical properties.

Appendix

In this appendix, we present an explicit development on how we transform Eq. (3) into Eq. (4); first let us consider Eq. (3),

$$R_\alpha = \frac{1}{\alpha - 1} \left(\ln \frac{1}{k_{nm}} \sum_{i=1}^n \sum_{j=1}^m p_i^\alpha q_j^{1-\alpha} + \ln \frac{1}{k_{nm}} \sum_{i=1}^n \sum_{j=1}^m p_i^{1-\alpha} q_j^\alpha \right). \quad (20)$$

When one applies the logarithm properties, $\ln(a) + \ln(b) = \ln(ab)$, then the term ab produces $n \times m$ operations; however, it is important to remark that the term $\ln(a) + \ln(b)$ produces $n + m$ operations. For this reason, it is necessary to introduce a constant $k_{nm} = nm$ in each term of Eq. (2) to avoid such inconsistency; in the following lines, we show how this constant, k_{nm} , works,

$$\begin{aligned} R_\alpha &= \frac{1}{\alpha - 1} \left(\ln \frac{1}{k_{nm}} \sum_{i=1}^n \sum_{j=1}^m p_i^\alpha q_j^{1-\alpha} + \ln \frac{1}{k_{nm}} \sum_{i=1}^n \sum_{j=1}^m p_i^{1-\alpha} q_j^\alpha \right), \\ &= \frac{1}{\alpha - 1} \ln \left[\left(\frac{1}{k_{nm}} \sum_{i=1}^n \sum_{j=1}^m p_i^\alpha q_j^{1-\alpha} \right) \left(\frac{1}{k_{nm}} \sum_{i=1}^n \sum_{j=1}^m p_i^{1-\alpha} q_j^\alpha \right) \right], \\ &= \frac{1}{\alpha - 1} \ln \left[\frac{1}{k_{nm}^2} \sum_{i=1}^n \sum_{j=1}^m \sum_{i=1}^n \sum_{j=1}^m p_i^\alpha q_j^{1-\alpha} p_i^{1-\alpha} q_j^\alpha \right], \\ &= \frac{1}{\alpha - 1} \ln \left[\frac{1}{k_{nm}^2} nm \sum_{i=1}^n \sum_{j=1}^m p_i q_j \right], \\ &= \frac{1}{\alpha - 1} \ln \sum_{i=1}^n \sum_{j=1}^m \frac{p_i q_j}{k_{nm}}, \\ &= \frac{1}{\alpha - 1} \ln \sum_{i=1}^n \sum_{j=1}^m p'_i q'_j. \end{aligned} \quad (21)$$

From this last equation, it is clear that,

$$\frac{1}{\alpha - 1} \ln \sum_{i=1}^n \sum_{j=1}^m p'_i q'_j = \frac{1}{\alpha - 1} \left(\ln \sum_{i=1}^n p'_i + \ln \sum_{j=1}^m q'_j \right). \quad (22)$$

Acknowledgements N. Flores-Gallegos wishes to thank the CONACyT, the PRODEP-SEP program for support.

Declarations

Conflict of interest The author declare that he has no conflict of interest.

References

- N. Nikolova, J. Jaworska, *QSAR Comb. Sci.* **22**, 1006 (2003)
- R. Carbo, L. Leyda, M. Arnau, *Int. J. Quantum Chem.* **17**, 1185 (1980)
- P. Willett, *Mol. Inf.* **33**(6), 403 (2014)
- G.M. Maggiora, V. Shanmugasundaram, in *Chemoinformatics and Computational Chemical Biology, Methods in Molecular Biology*, vol. 672, ed. by J. Bajorath (Springer, Berlin, 2011)
- R. Todeschini, V. Consonni, H. Xiang, J.D. Holliday, M. Buscema, P. Willett, *J. Chem. Inf. Model.* **52**, 2884 (2012)
- G. Maggiora, M. Vogt, D. Stumpfe, J. Bajorath, *J. Med. Chem.* **57**(8), 3186 (2014)
- C.E. Shannon, *Bell Syst. Tech. J.* **27**, 379 and 623 (1948)
- P. Ziesche, *Int. J. Quantum Chem.* **56**, 363 (1995)
- P. Gersdorf, W. John, J. Perdew, P. Ziesche, *Int. J. Quantum Chem.* **61**, 935 (1997)
- A. Grassi, *Int. J. Quantum Chem.* **108**, 774 (2008)
- A. Grassi, *Int. J. Quantum Chem.* **111**, 2390 (2011)
- Garry T. Smith, Hartmut L. Schmider, Vedene H. Smith Jr., *Phys. Rev. A.* **65**, 32508–1 (2002)
- Alex D. Gottlieb, Norbert J. Mauser, *Phys. Rev. Lett.* **95**, 123003–1 (2005)
- Luigi Delle Site, *Int. J. Quantum Chem.* **115**, 1396 (2015)
- N. Flores-Gallegos, *Chem. Phys. Lett.* **666**, 62 (2016)
- N. Flores-Gallegos, *Chem. Phys. Lett.* **692**, 61 (2018)
- Paul Ziesche, Vedene H. Smith Jr., Minhhuu Hô, Sven P. Rudin, Peter Gersdorf, Manfred Taut, *J. Chem. Phys.* **110**, 6135 (1999)
- A. Rényi, *Proc. Fourth Berkeley Symp. Math. Stat. Prob.* **1**, 547 (1961)
- J. Havrda, F. Charvát, *Kybernetika* **3**(1), 30 (1967)
- C. Tsallis, *J. Stat. Phys.* **52**(1–2), 479 (1988)
- J.C. Bolívar, A. Nagy, E. Romera, *Phys. A* **498**, 66 (2018)
- B. Godó, A. Nagy, *Chaos* **27**, 073104 (2017)
- B. Godó, A. Nagy, *Chaos* **26**, 083102 (2016)
- A. Nagy, E. Romera, *Europhys. Lett.* **109**, 60002 (2015)
- B. Godó, A. Nagy, *J. Phys. Conf. Ser.* **410**, 012090 (2013)
- E. Romera, R. del Real, M. Calixto, S. Nagy, A. Nagy, *J. Math. Chem.* **51**, 620 (2013)
- A. Nagy, E. Romera, *Phys. A* **391**, 3650 (2012)
- M. Calixto, A. Nagy, I. Paradela, E. Romera, *Phys. Rev. A.* **85**, 053813 (2012)
- I. Hornyák, A. Nagy, *Int. J. Quantum Chem.* **112**, 1285 (2012)
- E. Romera, A. Nagy, *Phys. Lett. A.* **375**, 3066 (2011)
- A. Nagy, E. Romera, *Phys. Lett. A.* **373**, 844 (2009)
- A. Nagy, E. Romera, *Int. J. Quantum Chem.* **109**, 2490 (2009)
- E. Romera, R. López-Ruiz, J. Sanudo, A. Nagy, *Int. Rev. Phys.* **3**, 207 (2009)
- E. Romera, A. Nagy, *Phys. Lett. A.* **372**, 4918 (2008)
- E. Romera, A. Nagy, *Phys. Lett. A.* **372**, 6823 (2008)
- N. Flores-Gallegos, *J. Math. Chem.* **59**(7), 1822 (2021)
- Tim van Erven, Peter Harremoës, *IEEE Trans. Inf. Theory* **60**(7), 3797 (2014)

38. Ramon Carbó-Dorca, Emili Besalú, J. Mol. Struct. (Theochem) **451**, 11 (1998)
39. P. Bultinck, X. Gironés, R. Carbó-Dorca, in *Reviews in Computational Chemistry*, vol. 21. ed. by K.B. Lipkowitz, R. Larter, T.R. Cundari (John Wiley & Sons, Inc., 2005), pp. 127–207
40. R. Carbó, B. Calabuig, J. Quantum Chem. **42**, 1681 (1992)
41. R. Carbó, B. Calabuig, J. Quantum Chem. **42**, 1696 (1992)
42. A. Bhattacharyya, News Bull. Calcutta Math. Soc. **35**, 99 (1943)
43. S. Kullback, R.A. Leibler, Ann. Math. Stat. **22**(1), 79 (1951)
44. P.A.M. Dirac, Math. Proc. Camb. Philos. Soc. **26**(3), 376 (1930)
45. R.G. Parr, W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, Oxford, 1989), pp. 87–104
46. Robert S. Mulliken, J. Chem. Phys. **2**, 782 (1934)
47. Shamus A. Blair, Ajit J. Thakkar, Comput. Theor. Chem. **1043**, 13 (2014)
48. Gaussian 09, Revision E.01, M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, O. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian, Inc., Wallingford CT (2013)
49. F. Soler-Rodríguez, M.P. Míguez-Santiyán, *Encyclopedia of Toxicology*, 3rd edn. (Academic Press, London, 2014), pp. 512–514

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.