Analysis of complexity measures and information planes of selected molecules in position and momentum spaces†

Rodolfo O. Esquivel,*[abc] Juan Carlos Angulo,[bc] Juan Antolín,[cd] Jesús S. Dehesa,[bc] Sheila López-Rosa[bc] and Nelson Flores-Gallegos[e]

Received 22nd December 2009, Accepted 18th March 2010
First published as an Advance Article on the web 14th May 2010
DOI: 10.1039/b927055h

The Fisher–Shannon and LMC shape complexities and the Shannon–disequilibrium, Fisher–Shannon and Fisher–disequilibrium information planes, which consist of two localization–delocalization factors, are computed in both position and momentum spaces for the one-particle densities of 90 selected molecules of various chemical types, at the CISD/6-311++G(3df,2p) level of theory. We found that while the two measures of complexity show general trends only, the localization–delocalization planes clearly exhibit chemically significant patterns. Several molecular properties (energy, ionization potential, total dipole moment, hardness, electrophilicity) are analyzed and used to interpret and understand the chemical nature of the composite information–theoretic measures above mentioned. Our results show that these measures detect not only randomness or localization but also pattern and organization.

1. Introduction

There has been a great interest in the last few years in applying complexity tools to study numerous chemical and biological phenomena. Complexity measures are understood as general indicators of pattern, structure, and correlation in systems or processes. Several alternative mathematical notions have been proposed for quantifying the concepts of complexity and information, including the algorithmic information theory of Kolmogorov and Chaitin,[1] the classical information theory of Shannon and Weaver,[2] Fisher information,[3] and the logical[4] and the thermodynamical[5] depths, among others. Some of them share rigorous connections with others as well as with Bayes and information theory.[6] The term complexity possesses many different meanings: algorithmic, geometrical, computational, stochastic, effective, statistical, and structural among others and it has been utilized in many fields: dynamical systems, disordered systems, spatial patterns, language, multielectronic systems, cellular automata, neuronal networks, self-organization, DNA analyses, social sciences, among others.[7,8]

Although there is no general agreement about the definition of complexity, its quantitative characterization has been an important subject of research and it has received considerable attention.[9,10] The characterization of complexity is not unique and the utility of each definition depends on the type of system or process, the level of the description, and the scale of the interactions among elementary particles, atoms, molecules, biological systems, *etc*. Fundamental concepts such as uncertainty or randomness are frequently employed in the definitions of complexity, although some other concepts such as clustering, order, localization or organization might be also important for characterizing the complexity of systems or processes.

It is not clear how the aforementioned concepts, such as *disorder* or *randomness*, might intervene in the definitions so as to quantitatively assess the complexity of the system. However, recent proposals have formulated this quantity as a product of two factors, taking into account *order*/*disequilibrium* and *disorder*/*uncertainty*, respectively. This is the case of the definition of López–Ruiz–Mancini–Calbet (LMC) shape complexity[9–12] that, like others, satisfies the boundary conditions by reaching its minimal value in the extreme ordered and disordered limits. The LMC complexity measure has been criticized,[11] modified[12] and generalized[13] leading to a useful estimator which satisfies several desirable properties of invariance under scaling, translation, and replication.[14,16] The utility of this improved complexity has been verified in many fields[8] and allows reliable detection of periodic, quasi-periodic, linear stochastic, and chaotic dynamics.[14–16] The LMC measure is constructed as the product of two important information-theoretic quantities (see below): the so-called disequilibrium D (also known as self-similarity[17] or information energy[18]), which quantifies the departure of the probability density from uniformity[15] and the Shannon entropy S, which

[a] *Departamento de Química, Universidad Autónoma Metropolitana-Iztapalapa, 09340-México D.F., México. E-mail: esquivel@xanum.uam.mx*
[b] *Departamento de Física Atómica, Molecular y Nuclear, Universidad de Granada, 18071-Granada, Spain*
[c] *Instituto Carlos I de Física Teórica y Computacional, Universidad de Granada, 18071-Granada, Spain*
[d] *Departamento de Física Aplicada, EUITIZ, Universidad de Zaragoza, 50018-Zaragoza, Spain*
[e] *Unidad Profesional Interdisciplinaria de Ingeniería, Campus Guanajuato del Instituto Politécnico Nacional, 36275-Guanajuato, México*

† Electronic supplementary information (ESI) available: Table TS1 (molecular properties), Table TS2 (complexity measures) and Table TS3 (information plane measures). See DOI: 10.1039/b927055h

is a general measure of randomness or uncertainty of the probability density.[2] Both global quantities are closely related to the measure of spread of a probability distribution.

On the other hand, the Fisher–Shannon product FS has been employed as a measure of atomic correlation[19] and more recently it has been defined as a statistical complexity measure,[20] in terms of the Shannon and Fisher information measures, so as to combine both, the global character (depending on the distribution as a whole) and the local one (in terms of the gradient of the distribution), respectively, while preserving the aforementioned desirable properties. The Fisher information I itself plays a fundamental role in different physical problems, such as the derivation of the non-relativistic quantum-mechanical equations by means of the minimum I principle, as well as the time-independent Kohn–Sham equations and the time-dependent Euler equation.[21] More recently, the Fisher information has been employed also as an intrinsic accuracy measure for specific atomic models and densities[22] as well as for general quantum-mechanical central potentials.[23] Also, the concept of phase-space Fisher information has been analyzed for hydrogen-like atoms and the isotropic harmonic oscillator,[24] where both position and momentum variables are included.

The FS complexity is defined in terms of the product of the Fisher information I, and the power entropy, J—explicitly defined in terms of the Shannon entropy (see below)—which is chosen to preserve the general complexity properties. As compared to the LMC complexity, aside of the explicit dependence on the Shannon entropy, the Fisher–Shannon complexity replaces the disequilibrium global factor D by the Fisher local one to quantify the level of *organization* of a given system. Several applications have been carried out, particularly the ones concerning with atomic distributions in position and momentum spaces, where the FS complexity is shown to provide relevant information on atomic shell structure and ionization processes.[20,25]

The goal of the present study is to perform an information–theoretical analysis by use of complexity measures in order to analyze and quantify the information content of a set of ninety molecular systems of different chemical type, as a probe for studying statistical complexity. Focus will be set on the recognition of patterns of *uncertainty*, *order* and *organization* by employing several molecular properties such as energy, ionization potential, hardness and electrophilicity. The above mentioned complexity measures and their associated informational planes will be analyzed in terms of the chemical properties, number of electrons and geometrical features.

The organization of the paper is as follows: in section 2 we defined the complexity measures along with their information-theoretic components and the chemical properties employed throughout the study. In section 3 we compute the information components as well as the Fisher–Shannon and LMC complexities. These information functionals of the one-particle density are computed in position (r) and momentum (p) spaces, as well as in a joint product space (rp) that contains more complete information about the system. Besides, the Fisher–Shannon (I–J) and the disequilibrium–Shannon (D–L) planes are studied to identify both, pattern and organization. In section 4, some conclusions are given.

2. Information-theoretic measures, complexities and chemical properties

In the independent-particle approximation, the total density distribution in a molecule is a sum of contribution from the electrons in each of the occupied orbitals. This is the case in both r-space and p-space, position and momentum respectively. In momentum space, the total electron density, $\gamma(\vec{p})$, is obtained through the molecular momentals (momentum-space orbitals) $\varphi_i(\vec{p})$, and similarly for the position-space density, $\rho(\vec{r})$, through the molecular position-space orbitals $\phi_i(\vec{r})$. The momentals can be obtained by three-dimensional Fourier transformation of the corresponding orbitals (and conversely) preserving their normalization.

$$\varphi_i(\mathbf{p}) = (2\pi)^{-3/2}\int \mathrm{d}\mathbf{r}\, \exp(-i\mathbf{p}\cdot\mathbf{r})\phi_i(\mathbf{r}) \tag{1}$$

Standard procedures for the Fourier transformation of position space orbitals generated by *ab initio* methods have been described.[26] The orbitals employed in *ab initio* methods are linear combinations of atomic basis functions and since analytic expressions are known for the Fourier transforms of such basis functions,[27] the transformations of the total molecular electronic wavefunction from position to momentum space is computationally straightforward.[28]

As we mentioned in the introduction, the LMC complexity is defined through the product of two relevant information–theoretic measures. So that, for a given probability density in position space, $\rho(\vec{r})$, the C(LMC) complexity is given by:

$$C_r(\mathrm{LMC}) = D_r \exp(S_r) = D_r L_r \tag{2}$$

where D_r is the disequilibrium

$$D_r = \int \rho^2(\mathbf{r})\, \mathrm{d}\mathbf{r} \tag{3}$$

and S is the Shannon entropy

$$S_r = -\int \rho(\mathbf{r}) \ln \rho(\mathbf{r})\, \mathrm{d}^3\mathbf{r} \tag{4}$$

from which the exponential entropy $L_r = \exp(S_r)$ is defined. Similar expressions for the LMC complexity measure in the conjugated momentum space might be defined for a distribution $\gamma(\vec{p})$

$$C_p(\mathrm{LMC}) = D_p \exp(S_r) = D_p L_p \tag{5}$$

It is important to mention that the LMC complexity of a system must comply with the following lower bound:[29]

$$C(\mathrm{LMC}) \geq 1 \tag{6}$$

The FS complexity in position space, C_r(FS), is defined in terms of the product of the Fisher information

$$I_r = \int \rho(\mathbf{r}) |\vec{\nabla} \ln \rho(\mathbf{r})|^2\, \mathrm{d}^3\mathbf{r} \tag{7}$$

and the power entropy in position space, J_r

$$J_r = \frac{1}{2\pi e} \exp(\tfrac{2}{3} S_r), \tag{8}$$

which depends on the Shannon entropy defined above. So that, the FS complexity in position space is given by

$$C_r(\mathrm{FS}) = I_r \cdot J_r \tag{9}$$

and similarly

$$C_p(\mathrm{FS}) = I_p \cdot J_p \tag{10}$$

in momentum space.

Let us remark that the factors in the power Shannon entropy J are chosen to preserve the invariance under scaling transformations, as well as the rigorous lower bound.[30]

$$C(\mathrm{FS}) \geq n \tag{11}$$

with n being the space dimensionality, thus providing a universal lower bound to FS complexity. The definition in eqn (8) corresponds to the particular case $n = 3$, the exponent containing a factor $2/n$ for arbitrary dimensionality.

It is worthwhile noting that the aforementioned inequalities remain valid for distributions normalized to unity, which is the choice that it is employed throughout this work for the three-dimensional molecular case.

Aside of the analysis of the position and momentum information measures, we have considered it useful to study these magnitudes in the product space, *rp*-space, characterized by the probability density $f(\vec{r}, \vec{p}) = \rho(\vec{r})\gamma(\vec{p})$, where the complexity measures are defined as

$$C_{rp}(\mathrm{LMC}) = D_{rp}L_{rp} = C_r(\mathrm{LMC})C_p(\mathrm{LMC}), \tag{12}$$

and

$$C_{rp}(\mathrm{FS}) = I_{rp}J_{rp} = 2\pi\mathrm{e}C_r(\mathrm{FS})C_p(\mathrm{FS}), \tag{13}$$

where last equality arises from the definition of J_{rp} as the power entropy of $S_{rp} = S_r + S_p$. From the above two equations, it is clear that the features and patterns of both LMC and FS complexity measures in the product space will be determined by those of each conjugated space. However, the numerical analyses carried out in the next section, reveal that the momentum space contribution plays a more relevant role as compared to the one in position space.

With the purpose of organizing and characterizing the complexity features of the molecular systems under study, we have computed several reactivity properties such as the ionization potential (IP), the total dipole moment, the hardness (η) and the electrophilicity index (ω). These properties were obtained at the Hartree–Fock level of theory (HF) in order to employ Koopmans' theorem,[31] for relating the first vertical ionization energy and the electron affinity to the HOMO and LUMO energies, which are necessary to calculate the conceptual DFT properties. The hardness is obtained within this framework[32] through

$$\eta = \frac{1}{2S} \approx \frac{\varepsilon_{\mathrm{LUMO}} - \varepsilon_{\mathrm{HOMO}}}{2} \tag{14}$$

where ε denotes the frontier molecular orbital energies and S stands for the softness of the system. It is worth mentioning that the factor 1/2 in eqn (14) was put originally to make the hardness definition symmetrical with respect to the chemical potential[33]

$$\mu = \left(\frac{\partial E}{\partial N}\right)_{\nu(r)} = \frac{\varepsilon_{\mathrm{LUMO}} + \varepsilon_{\mathrm{HOMO}}}{2} \tag{15}$$

although it has been recently disowned.[34] In general terms, the chemical hardness and softness are good descriptors of chemical reactivity. The former has been employed[34,35] as a measure of the reactivity of a molecule in the sense of the resistance to changes in the electron distribution of the system, *i.e.*, molecules with larger values of η are interpreted as being the least reactive ones. In contrast, the S index quantifies the polarizability of the molecule[36] and hence soft molecules are more polarizable and possess predisposition to acquire additional electronic charge.[37] The chemical hardness η is a central quantity for use in the study of reactivity through the hard and soft acids and bases principle.[38]

The electrophilicity index,[39] ω, allows a quantitative classification of the global electrophilic nature of a molecule within a relative scale. Electrophilicity index of a system in terms of its chemical potential and hardness is given by the expression

$$\omega = \frac{\mu^2}{2\eta} \tag{16}$$

The electrophilicity is also a good descriptor of chemical reactivity, which quantifies the global electrophilic power of the molecules[32] (predisposition to acquire an additional electronic charge).

3. LMC and FS complexity measures of molecular systems

The molecular set chosen for the study includes different types of chemical organic and inorganic systems (aliphatic and aromatic hydrocarbons, alcohols, ethers, ketones). The set represents a variety of closed shell systems, radicals, isomers as well as molecules with heavy atoms such as sulphur, chlorine, magnesium and phosphorous. The geometries needed for the single point energy calculations above referred were obtained from the *Computational Chemistry Comparison and Benchmark DataBase* from NIST.[40] The molecular set might be organized by isoelectronic groups as follows:

$N = 10$: NH_3 (ammonia)

$N = 12$: LiOH (lithium hydroxide)

$N = 14$: HBO (boron hydride oxide), Li_2O (dilithium oxide)

$N = 15$: HCO (formyl radical), NO (nitric oxide)

$N = 16$: H_2CO (formaldehyde), NHO (nitrosyl hydride), O_2 (oxygen)

$N = 17$: CH_3O (methoxy radical)

$N = 18$: CH_3NH_2 (methyl amine), CH_3OH (methyl alcohol), H_2O_2 (hydrogen peroxide), NH_2OH (hydroxylamine)

$N = 20$: NaOH (sodium hydroxide)

$N = 21$: BO_2 (boron dioxide), C_3H_3 (radical propargyl), MgOH (magnesium hydroxide), HCCO (ketenyl radical)

$N = 22$: C_3H_4 (cyclopropene), CH_2CCH_2 (allene), CH_3CCH (propyne), CH_2NN (diazomethane), CH_2CO (ketene), CH_3CN (acetonitrile), CH_3NC (methyl isocyanide), CO_2 (carbon dioxide), FCN (cyanogen fluoride), HBS (hydrogen boron sulfide), HCCOH (ethynol), HCNO (fulminic acid), HN_3 (hydrogen azide), HNCO (isocyanic acid), HOCN (cyanic acid), N_2O (nitrous oxide), NH_2CN (cyanamide)

$N = 23$: NO_2 (nitrogen dioxide), NS (mononitrogen monosulfide), PO (phosphorus monoxide), C_3H_5 (allyl radical), CH_3CO (acetyl radical)

N = 24: C_2H_4O (ethylene oxide), C_2H_5N (aziridine), C_3H_6 (cyclopropane), CF_2 (difluoromethylene), CH_2O_2 (dioxirane), CH_3CHO (acetaldehyde), $CHONH_2$ (formamide), FNO (nitrosyl fluoride), H_2CS (thioformaldehyde), HCOOH (formic acid), HNO_2 (nitrous acid) $NHCHNH_2$ (aminomethanimine), O_3 (ozone), SO (sulfur monoxide)

N = 25: $CH_2CH_2CH_3$ (*n*-propyl radical), CH_3CHCH_3 (isopropyl radical), CH_3OO (methylperoxy radical), FO_2 (dioxygen monofluoride), NF_2 (difluoroamine radical), CH_3CHOH (ethoxy radical), CH_3S (thiomethoxy)

N = 26: C_3H_8 (propane), $CH_3CH_2NH_2$ (ethylamine), CH_3CH_2OH (ethanol), CH_3NHCH_3 (dimethylamine), CH_3OCH_3 (dimethyl ether), CH_3OOH (methyl peroxide), F_2O (difluorine monoxide)

N = 30: ClCN (chlorocyanogen), OCS (carbonyl sulfide), SiO_2 (silicon dioxide)

N = 31: PO_2 (phosphorus dioxide), PS (phosphorus sulfide)

N = 32: ClNO (nitrosyl chloride), S_2 (sulfur diatomic), SO_2 (sulfur dioxide)

N = 33: OClO (chlorine dioxide), ClO_2 (chlorine dioxide)

N = 34: CH_3CH_2SH (ethanethiol), CH_3SCH_3 (dimethyl sulfide), H_2S_2 (hydrogen sulfide), SF_2 (sulfur difluoride)

N = 38: CS_2 (carbon disulfide)

N = 40: CCl_2 (dichloromethylene), S_2O (disulfur monoxide)

N = 46: $MgCl_2$ (magnesium dichloride)

N = 48: S_3 (sulfur trimer), $SiCl_2$ (dichlorosilylene)

N = 49: ClS_2 (sulfur chloride)

The electronic structure calculations performed in the present study for the whole set of molecules were obtained with the Gaussian 03 suite of programs[41] at the CISD/6-311++G(3df,2p) level of theory. For this set of molecules we have calculated all information and complexity measures defined in the previous section, *i.e.*, S, D, I, J, C(LMC), C(FS), in both position and momentum spaces as well as in the product space by employing software developed in our laboratory along with 3D numerical integration routines[42] and the DGRID suite of programs.[28] As mentioned above, the values of the conceptual DFT properties have been obtained at the HF/6-311++G(3df,2p) level of theory. All calculated quantities in this study are given in atomic units, and their values can be consulted in the ESI†.

3.1 Complexity measures

In contrast with the atomic case, where the complexities possess a high level of natural organization provided by periodicity properties,[20] the molecular case requires some sort of organization/classification which could be affected by many factors (structural, energetic, entropic, *etc.*). Thus we have analyzed the molecular complexities, C(LMC) and C(FS), as functions of the main chemical properties of interest, *i.e.*, the total energy, the dipole moment, the ionization potential, the hardness and the electrophilicity, establishing a link between the different complexity measures and the chemical properties so as to provide an insight into their *organization*, *order* and *uncertainty* features.

In Fig. 1, we have plotted the C(FS) and C(LMC) as functions of the total energy of the molecules in the product space (*rp*). Firstly, it may be observed from this figure that both complexity measures possess a similar behavior. Secondly, a general trend is observed in that molecules with higher energy correspond to higher complexity values, for both C(LMC) and C(FS), as compared to those which possess lower energies. Note that, according to the set of molecules studied in this work, they are grouped together according to four energy intervals: $E > -400$, $E \in [-700, -400]$, $E \in [-1000, -700)$, and $E < -1000$. In this figure we have indicated those molecules which correspond to the maximum and minimum values of C_{rp}(FS) within each group, noting that these molecules possess also maximum C_{rp}(LMC) values. It is worth noting that the maximum values correspond with molecules that contain one heavy atom at least and that the minimum complexities correspond with molecules having similar chemical geometry. In view of the results shown in this figure, it seems that the molecular complexity is also affected by several other chemical factors, such as the molecular structure (*e.g.*, lower complexity values in each group, energy-wise, corresponds with molecules of similar geometry), composition (*e.g.*, higher complexity corresponds with molecules containing heavy atoms), chemical functionality, reactivity, *etc.* Note that each complexity consists of two factors, one of them always defined in terms of the Shannon entropy S, whereas the other characterizes more specifically the corresponding complexity measure in terms of a global quantity (disequilibrium D) for C_{rp}(LMC) and a local one (Fisher information I) for C_{rp}(FS). Nevertheless, there are no relevant structural differences between both complexities, either based on D or I.

In Fig. 2 we have plotted the complexity values for the C_{rp}(LMC) and C_{rp}(FS) measures as a function of the number of electrons in the product space (note that both complexity measures are in a double-X axis graph). We may observe from Fig. 2 that both complexity measures behave in a similar fashion, *i.e.*, molecules with low number of electrons ($N < 26$) possess low complexities whereas molecules with larger number of electrons ($N > 26$) possess larger complexity values. A few exceptions may be noted from Fig. 2, *i.e.*, molecules with low number of electrons and higher complexities correspond to

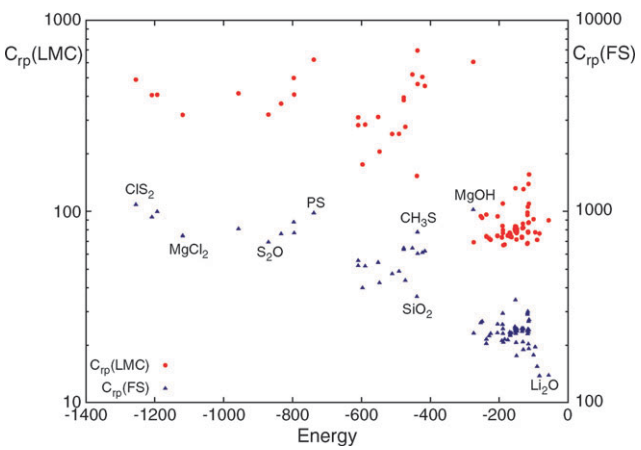

Fig. 1 C(LMC) (red circles) and C(FS) (blue triangles) complexities as a function of the total energy (a.u.) for the set of ninety molecules in the product space (*rp*).

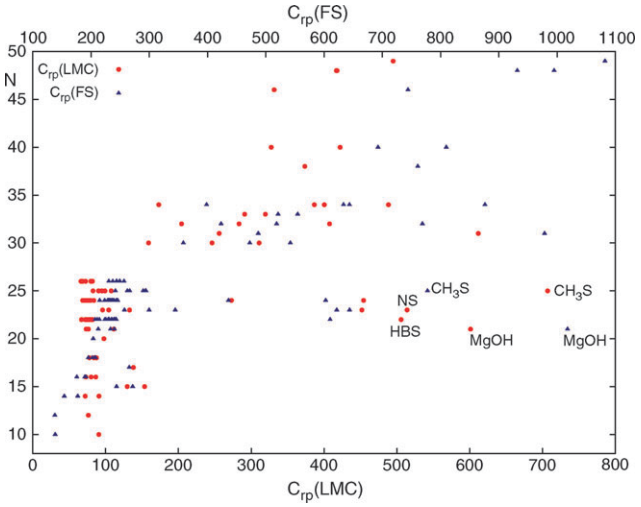

Fig. 2 *C*(LMC) (red circles) and *C*(FS) (blue triangles) complexities as a function of the number of electrons (*N*) for the set of ninety molecules in the product space (*rp*).

those containing heavier atoms: CH_3S, MgOH, NS and HBS for the C_{rp}(LMC), and MgOH and CH_3S for the C_{rp}(FS), as we have observed before in terms of the analysis based on the total energy. It is worth mentioning, similarly to the case of the results in Fig. 1, that several other factors may affect the molecular complexities as it was discussed above.

In order to analyze the influence of the chemical reactivity on the complexities of the set of studied molecules, we have plotted in Fig. 3 the hardness values *vs.* the LMC and FS complexities in the product space (*rp*) for the studied set of molecules. The general observations are that the LMC and FS complexities behave in the same way, both indicating a clear relationship with the hardness and hence with the chemical reactivity of the molecules. Besides, it might be observed that as the hardness increases, the complexity values decreases. This fact illustrates that molecules that are more stable chemically (resistance to changes in the electron distribution) possess low complexity values. Thus, the chemical reactivity seems to be directly related to complexity in that higher *C*(LMC) and *C*(FS) correspond with more reactive molecules, with very few exceptions which again correspond to systems with heavier atoms as we mentioned before (CH_3S, PS, MgOH for *C*(LMC) and ClS_2, MgOH, S_3 for *C*(FS)). It is worth mentioning that a similar analysis for the total dipole moment (in Table TS1, ESI†) might be performed and so we noted that molecules with higher complexity possess lower values for the dipole moment, *i.e.* those that are more polarizable and hence the most reactive (which in this case, correspond to the molecules containing heavier atoms).

The ionization potential (IP) is now employed as an indicator of the chemical stability of the molecules in relation with their complexities. In Fig. 4 it may be observed that molecules with higher IP values (more stable ones) are located at the right side of the figure, noting that stability is related with the molecular complexities in that higher LMC and FS complexities correspond with more reactive molecules (which are less stable). The electrophilicity index is a useful indicator of chemical reactivity quantifying the global electrophilic power of the molecules,

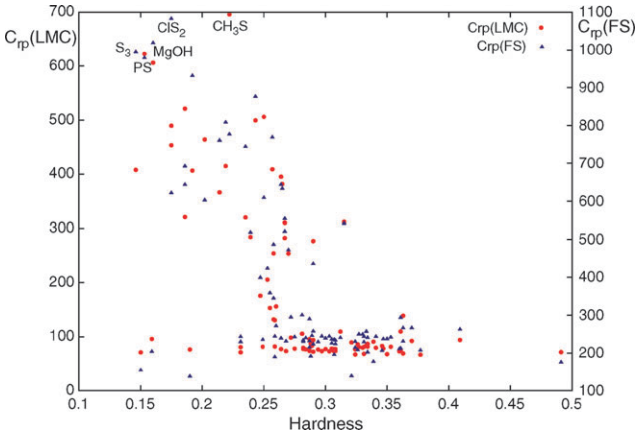

Fig. 3 *C*(LMC) (red circles) and *C*(FS) (blue triangles) complexities as a function of the hardness (a.u.) for the set of ninety molecules in the product space (*rp*).

as mentioned above. Thus, we have found useful to study the complexities, *C*(LMC) and *C*(FS), as a function of the electrophilicity in the product space (see Table TS1, ESI†). In contrast with the other analyzed properties, the electrophilicity (affected by both, the hardness and the chemical potential, eqn (16)), displays a more complicated behavior in that molecules with lower values of complexity are associated with molecules possessing lower electrophilicity, except for the NaOH molecule which is highly electrophilic.

It is noteworthy that all conclusions obtained from the analysis performed in the product space (*rp*) remain valid also when considering the momentum space (*p*) alone. Nevertheless, most of the structural features observed in the figures are better displayed in the product than in the momentum space, as a consequence of the joint effect arising when dealing simultaneously with the position variable, essentially by shifting away consecutive curves.

3.2 Information planes

In the search of pattern and organization we have found useful to analyze the set of studied molecules, through features such as their energy and the number of electrons, by plotting the

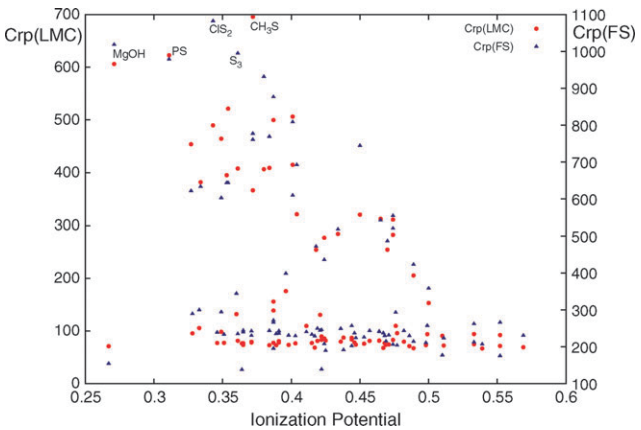

Fig. 4 *C*(LMC) (red circles) and *C*(FS) (blue triangles) complexities as a function of the ionization potential (a.u.) for the set of ninety molecules in the product space (*rp*).

contribution of each one of the information measures D (*order*) and L (*uncertainty*) to the total LMC complexity, and similarly with I (*organization*) and J (*uncertainty*) to the FS complexity. Thus, in Figs. 5 and 6 we have analyzed the behavior of the energy in the D–L_p and I_p–J_p planes, and in Fig. 7 and 8 the effect of the number of electrons in the D_r–L_r and I_p–J_p planes is displayed. For the energetic analysis of the information planes, we have found more useful to depict the corresponding ones to the momentum space, since the momentum density is directly related to the energy. In Fig. 5, we have plotted (in a double-logarithmic scale) the D_p–L_p plane for the set of molecules which are grouped together and labeled according to the energy intervals observed in Fig. 1, *i.e.*, E_{400} for molecules with $E > -400$ a.u., E_{700} for $E \in [-700, -400]$ a.u., E_{1000} for $E \in [-1000, -700)$ a.u., and E_{1400} for molecules with $E < -1000$ a.u. From Fig. 5 it is observed that the D–L plane is clearly separated into two regions, according to the $D{\cdot}L \geq 1$ inequality[29] (valid for position, momentum as well as product spaces), and the region below the line (equality) corresponds with the forbidden region. Parallel lines to this bound represent isocomplexity lines, showing that an increase (decrease) in *uncertainty*, L_p, along them is compensated by a proportional decrease (increase) of *order* (disequilibrium D_p), and higher deviations from this frontier are associated to greater LMC complexities. The general observation from this set of molecules is that groups with different energies are somewhat separated into different regions, *i.e.*, molecules with higher energies possess the highest values of J_p (more *uncertainty*), whereas for the disequilibrium values these are distributed over a wider range of values for all the groups, energy-wise. So it seems that the energy is related to the *uncertainty* of the systems either for L_p in Fig. 5 and J_p in Fig. 6 (discussed below). An interesting feature that is noteworthy from this figure is that low energy molecules behave more linearly and locate closer to the bound than higher energy molecules, *i.e.*, more energetic molecules seem to

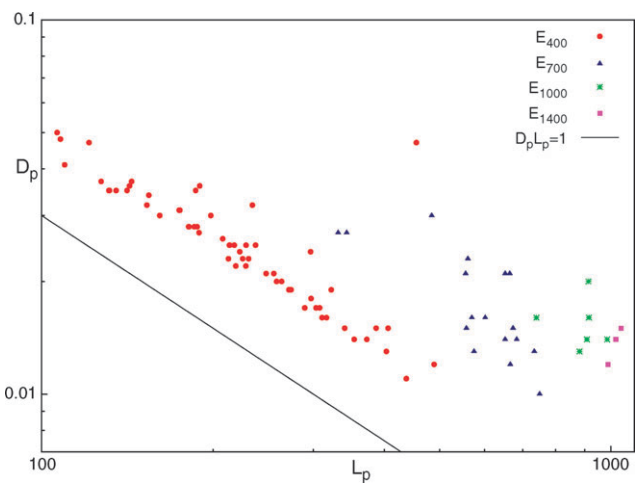

Fig. 5 Disequilibrium–Shannon plane (D–L) in momentum space for energetically differents groups: E_{400} for molecules with $E > -400$ (red circles), E_{700} for $E \in [-700, -400]$ (blue triangles), E_{1000} for $E \in [-1000, -700)$ (green stars), and E_{1400} for molecules with $E < -1000$ (magenta box). Double-logarithmic scale. Lower bound ($D_p{\cdot}L_P = 1$) is depicted by the black line.

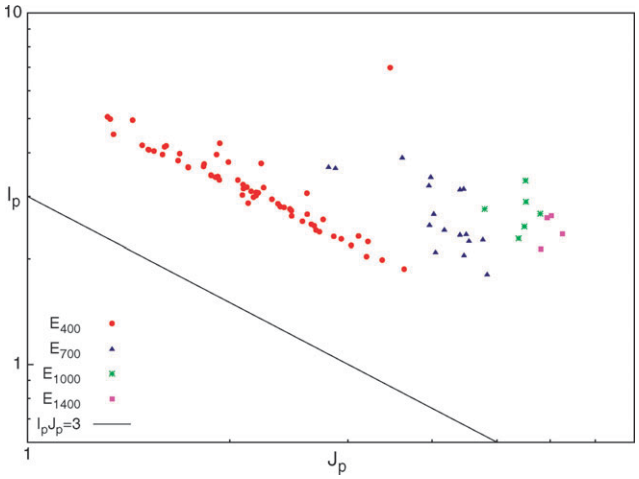

Fig. 6 Fisher–Shannon plane (I–J) in momentum space for energetically differents groups: E_{400} for molecules with $E > -400$ (red circles), E_{700} for $E \in [-700, -400]$ (blue triangles), E_{1000} for $E \in [-1000, -700)$ (green stars), and E_{1400} for molecules with $E < -1000$ (magenta box). Double-logarithmic scale. Lower bound ($I_p{\cdot}J_p = 3$) is depicted by the black line.

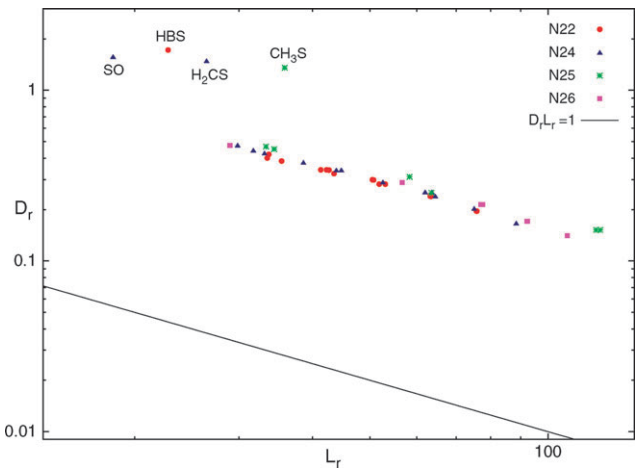

Fig. 7 Disequilibrium–Shannon plane (D–L) in position space of the isoelectronic series of 22 (red circles), 24 (blue triangles), 25 (green stars), and 26 (magenta box) electrons. Double-logarithmic scale. Lower bound ($D_r{\cdot}L_r = 1$) is depicted by the black line. Molecules with larger energy values are shown at the upper left corner of the Figure.

deviate from the isocomplexity lines. This observation deserves a deeper study with a larger number of molecules and with a wider range of energies.

In Fig. 6 we have plotted (in a double-logarithmic scale) the I_p–J_p plane for the same set of groups, energy-wise. At this point it is worth mentioning that there is a rigorous lower bound to the associated FS complexity, given in eqn (11), which is $C(\mathrm{FS}) = I{\cdot}J \geq$ constant (the constant being 3 for the conjugated spaces and $18\pi e$ for the product space as it was explained above in relation with eqn (13)). Fig. 6 indicates a division of the I_p–J_p plane into two regions where the straight line $I{\cdot}J =$ constant (drawn in the plane) divides it into an "allowed" (upper) and a "forbidden" (lower) part. Similar observations as those discussed in Fig. 5 apply also in this case,

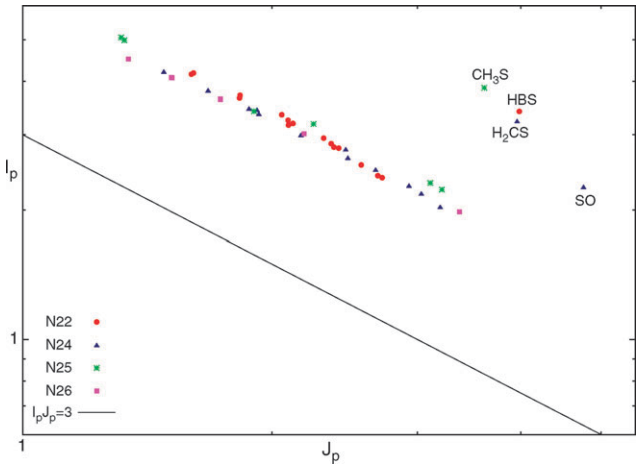

Fig. 8 Fisher–Shannon plane (I–J) in momentum space of the isoelectronic series of 22 (red circles), 24 (blue triangles), 25 (green stars), and 26 (magenta box) electrons. Double-logarithmic scale. Lower bound ($I_p \cdot J_p = 3$) is depicted by the black line. Molecules with larger energy values are shown at the upper left corner of the Figure.

i.e. those groups with larger energy values possess larger *uncertainty* as measured by the J_p values, whereas for *organization*, as measured through I_p, a wide spread is observed over the studied range of values. A similar conjecture as the one discussed above for the D_p–L_p plane might apply in this case in that lower energy molecules seems to obey a linear behavior (isocomplexity lines).

Our analysis continues with the study of the information planes for the isoelectronic series. In Fig. 7 the components of the LMC complexity in position space are depicted (in a double-logarithmic scale), for the series with $N = 22$, 24, 25 and 26 electrons, denoted by $N22$, $N24$, $N25$ and $N26$ in the figure, respectively, in the D_r–L_r plane. In this case each isoelectronic series follows a trajectory which shows a linear behaviour (similar trends are observed in momentum space D_p–L_p) with correlation coefficients close to one: $N22$ (0.989), $N24$ (0.991), $N25$ (0.989), $N26$ (0.981). Systems that are not in the isocomplexity lines belong to molecules with higher complexity values (in Table TS2, ESI†) which possess heavier atoms. This behavior means that in position space, higher complexity is due to higher disequilibrium (higher *order*) and lower *uncertainty* L_r. It is interesting to mention that the opposite behavior is observed in momentum space (not depicted), *i.e.*, higher complexity values correspond with lower disequilibrium and higher *uncertainty*. It is also worthy to note that all isocomplexity lines, representing the isoelectronic molecular series with $N = 22$, 24, 25 and 26 electrons, show large deviations (higher LMC complexities) from the rigorous lower bound as it may be observed from Fig. 7 in position space.

Proceeding with the analysis of *pattern* and *organization* for the isoelectronic series, we have analyzed the contribution of each one of the information measures I and J to the total FS complexity. This is done in Fig. 8 for some of the isoelectronic molecular series with $N = 22$, 24, 25 and 26 electrons in the momentum space through the information I_p–J_p plane. Parallel lines to the constant represent isocomplexity lines, and higher deviations from this frontier are associated with greater FS complexities. Over these lines, an increase (decrease) in *uncertainty* (J) gets balanced by a proportional decrease (increase) of *accuracy* (I). Such a parallel shape is displayed by all isoelectronic series in momentum space, as shown in Fig. 8, and we have verified their linear behaviour by a linear regression analysis with the following correlation coefficients: $N22$ (0.989), $N24$ (0.994), $N25$ (0.993), $N26$ (0.998). Notice that systems that are not in the isocomplexity lines belong to higher complexity molecules as we have previously discussed. They possess heavier atoms, as observed from Fig. 8, possessing higher values of J_p (more *uncertainty*) which provokes their higher complexity. It is also worth noting that all isocomplexity lines representing the same isoelectronic molecular series show large deviations (higher FS complexities) from the rigorous lower bound $I \cdot J \geq 3$ as it may be observed from the figure. On the other hand, in the conjugated position space I_r–J_r (not depicted), we may observe a similar trend, *i.e.*, each isoelectronic series possess a linear behaviour except for the molecules with highest complexity values which are not on the isocomplexity lines, with lower values of J_r and higher values of I_r in contrast with Fig. 8.

Notwithstanding that not all information products are good candidates to form complexity measures, *i.e.*, preserving the desirable properties of invariance under scaling, translation and replication, we have found interesting to study the plane I–D, with the purpose of analyzing patterns of *order–organization*. Note that this product fails to be invariant under scale transformation.[14] Thus, in Fig. 9 and 10 we have plotted the I–D planes for the set of groups, energy-wise, studied above, and for some of the isoelectronic series with $N = 22$, 24, 25 and 26 electrons, respectively.

It can be observed from Fig. 9 that there exists a relationship between *order* (disequilibrium) and *organization* (Fisher information) for the set of studied molecules in momentum space within the I_p–D_p plane, *i.e.*, that molecules with higher values of D_p possess higher I_p values. It is apparent from the figure that a linear relationship between I and D is obeyed for

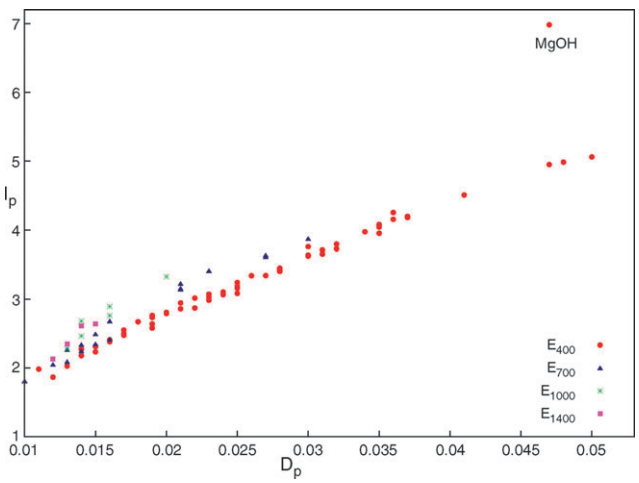

Fig. 9 Fisher–Disequilibrium plane (I–J) in momentum space for energetically different groups: E_{400} for molecules with $E > -400$ (red circles), E_{700} for $E \in [-700, -400]$ (blue triangles), E_{1000} for $E \in [-1000, -700)$ (green stars), and E_{1400} for molecules with $E < -1000$ (magenta box). Double-logarithmic scale.

Table 1 Chemical properties and complexity measures for the isomers HCNO, HNCO and HOCN in atomic units (a.u.)

Molecule	Energy	Ionization potential	Hardness	Electrophilicity	C_{rp}(LMC)	C_{rp}(FS)
HCNO	−168.134	0.403	0.294	0.020	76.453	229.303
HNCO	−168.261	0.447	0.308	0.031	74.159	223.889
HOCN	−168.223	0.453	0.305	0.036	75.922	225.656

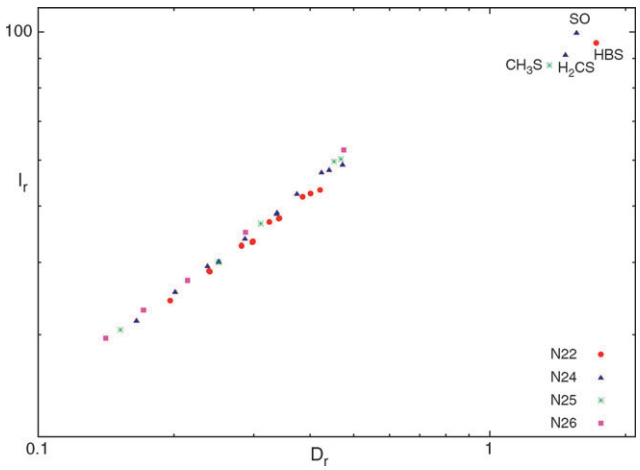

Fig. 10 Fisher–Disequilibrium plane (*I*–*D*) in position space of the isoelectronic series of 22 (red circles), 24 (blue triangles), 25 (green stars), and 26 (magenta box) electrons. Double-logarithmic scale.

these groups of energetically similar molecules. This linear behavior is observed for all molecules of the lower energetic group except for MgOH which is one of the higher complexity ones in terms of *C*(LMC) and *C*(FS). To the best of our knowledge, this apparent linear behavior between *D* and *I* has not been studied before.

A similar linear behavior might be observed from Fig. 10 in the corresponding I_r–D_r plane, showing a positive slope for all the isoelectronic molecular series, which means that as the molecular *order* increases (higher *D*) their *organization* also increases (higher *I*). Interestingly, Fig. 10 shows that the *I*–*D* plane is useful to detect molecular patterns of *order–organization* except for molecules of higher complexity (SO, HBS, H_2CS, CH_3S) which do not fit with such a linear description.

Finally, it is illustrative to analyze the particular case of three isoelectronic isomers: HCNO (fulminic acid), HNCO (isocyanic acid) and HOCN (cyanic acid) in order to analyze their chemical properties with respect to their complexity values. From an experimental side it is known that cyanic and isocyanic acids are isomers of fulminic acid (H–C═N–O) which is an unstable compound.[43] From the Table 1 we may corroborate that this is indeed the case in that fulminic acid possesses the lowest ionization potential (less stability) but larger values for the complexity measures. According to our discussion above for the chemical properties, this is indeed a more reactive molecule (lowest hardness value).

Conclusions

We have investigated the internal disorder of 90 molecules by means of five composite information–theoretic measures: the Fisher–Shannon and LMC shape complexities and three planes of information. The study of these measures in both position and momentum spaces is required in order to obtain a more complete description of the information–theoretical interpretation of the molecular systems.

According to the analysis of the LMC and FS complexities a few general trends are to be noted in that molecules that show higher complexity values correspond with molecules possessing higher energies, larger number of electrons, smaller hardness values (more chemically reactive) and smaller values for the ionization potential (less stable). However, our study reveals that there is no clear correlation between molecular complexity and chemical reactivity suggesting that the former is affected altogether by several chemical factors, such as the molecular structure (*e.g.*, molecules with lower complexity possess similar geometry), composition (*e.g.*, higher complexity correspond with molecules containing heavier atoms), chemical functionality, reactivity, *etc*.

The analysis of the information planes reveals that the molecular energies are related to the uncertainty of the systems as measured by L_p or J_p. Besides, we have observed that low energy molecules behave more linearly and closer to the lower bounds than higher energy molecules, *i.e.*, more energetic molecules seem to deviate from the isocomplexity lines. The results of this study indicate that further investigations along the lines of analyzing a larger number of molecules with a wider range of energies are necessary in order to improve our understanding of molecular complexity.

Acknowledgements

We wish to thank José María Pérez-Jordá and Miroslav Kohout for kindly providing with their numerical codes. R. O. E. wishes to thank Juan Carlos Angulo and Jesús Sánchez-Dehesa for their kind hospitality during his sabbatical stay on the Departamento de Física Atómica, Molecular y Nuclear at the Universidad de Granada, Spain. We acknowledge financial support through Mexican grants 08226 CONACyT, PIFI PROMEP-SEP and Spanish grants MICINN projects FIS-2008-02380, FQM-4643 and P06-FQM-2445 of Junta de Andalucía. J. S. D., J. C. A., S. L. R., R. E. O. and J. A. belong to the Andalusian research group FQM-0207. Allocation of supercomputing time from the Sección de Supercomputación at CSIRC-Universidad de Granada, the Laboratorio de Supercómputo y Visualización at UAM and to the Departamento de Supercómputo at DGSCA-UNAM is gratefully acknowledged. Finally, we would like to thank the referees whose comments helped to enrich the paper.

Notes and references

1 A. N. Kolmogorov, *Probl. Inf. Transm.*, 1965, **1**, 3; G. Chaitin, *J. ACM*, 1966, **13**, 547.
2 C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1949.

3 R. A. Fisher, *Math. Proc. Cambridge Philos. Soc.*, 1925, **22**, 700; B. R. Frieden, *Science from Fisher Information*, Cambridge University Press, Cambridge, 2004.
4 C. H. Bennet, in *The Universal Turing Machine: a Half Century*, ed. R. Herhen, Oxford University Press, Oxford, 1988, pp. 227–257.
5 S. Lloyd and H. Pagels, *Ann. Phys.*, 1988, **188**, 186.
6 P. M. B. Vitanyi and M. Li, *IEEE Trans. Inf. Theory*, 2000, **46**, 446.
7 C. R. Shalizi, K. L. Shalizi and R. Haslinger, *Phys. Rev. Lett.*, 2004, **93**, 118701.
8 O. A. Rosso, M. T. Martin and A. Plastino, *Phys. A*, 2003, **320**, 497; K. Ch. Chatzisavvas, Ch. C. Moustakidis and C. P. Panos, *J. Chem. Phys.*, 2005, **123**, 174111; A. Borgoo, F. De Proft, P. Geerlings and K. D. Sen, *Chem. Phys. Lett.*, 2007, **444**, 186.
9 D. P. Feldman and J. P. Crutchfield, *Phys. Lett. A*, 1998, **238**, 244.
10 P. W. Lamberti, M. T. Martin, A. Plastino and O. A. Rosso, *Phys. A*, 2004, **334**, 119.
11 C. Anteonodo and A. Plastino, *Phys. Lett. A*, 1996, **223**, 348.
12 R. G. Catalán, J. Garay and R. López-Ruiz, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2002, **66**, 011102; M. T. Martin, A. Plastino and O. A. Rosso, *Phys. Lett. A*, 2003, **311**, 126.
13 R. López-Ruiz, *Biophys. Chem.*, 2005, **115**, 215.
14 T. Yamano, *J. Math. Phys.*, 2004, **45**, 1974.
15 R. López-Ruiz, H. L. Mancini and X. Calbet, *Phys. Lett. A*, 1995, **209**, 321.
16 T. Yamano, *Phys. A*, 2004, **340**, 131.
17 R. Carbó-Dorca, J. Arnau and L. Leyda, *Int. J. Quantum Chem.*, 1980, **17**, 1185.
18 O. Onicescu, *C. R. Acad. Sci. Paris A*, 1966, **263**, 25.
19 E. Romera and J. S. Dehesa, *J. Chem. Phys.*, 2004, **120**, 8906.
20 J. C. Angulo, J. Antolın and K. D. Sen, *Phys. Lett. A*, 2008, **372**, 670; K. D. Sen, J. Antolín and J. C. Angulo, *Phys. Rev. A: At., Mol., Opt. Phys.*, 2007, **76**, 032502.
21 A. Nagy, *J. Chem. Phys.*, 2003, **119**, 9401; R. Nalewajski, *Chem. Phys. Lett.*, 2003, **372**, 28.
22 A. Nagy and K. D. Sen, *Phys. Lett. A*, 2006, **360**, 291; K. D. Sen, C. P. Panos, K. Ch. Chatzisavvas and Ch. Moustakidis, *Phys. Lett. A*, 2007, **364**, 286.
23 E. Romera, P. Sánchez-Moreno and J. S. Dehesa, *J. Math. Phys.*, 2006, **47**, 103504; J. S. Dehesa, R. González-Férez and P. Sánchez-Moreno, *J. Phys. A: Math. Theor.*, 2007, **40**, 1845.
24 I. Hornyák and A. Nagy, *Chem. Phys. Lett.*, 2007, **437**, 132.
25 J. C. Angulo and J. Antolín, *J. Chem. Phys.*, 2008, **128**, 164109; J. Antolín and J. C. Angulo, *Int. J. Quantum Chem.*, 2009, **109**, 586.
26 D. C. Rawlings and E. R. Davidson, *J. Phys. Chem.*, 1985, **89**, 969.
27 P. Kaijser and V. H. Smith, Jr., *Adv. Quantum Chem.*, 1977, **10**, 37.
28 M. Kohout, program DGRID, version 4.2, 2007.
29 S. López-Rosa, J. C. Angulo and J. Antolín, *Phys. A*, 2009, **388**, 2081.
30 Dembo, T. A. Cover and J. A. Thomas, *IEEE Trans. Inf. Theory*, 1991, **37**, 1501.
31 T. A. Koopmans, *Physica*, 1934, **1**, 104; J. F. Janak, *Phys. Rev. B: Condens. Matter*, 1978, **18**, 7165.
32 R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, New York, 1989.
33 R. G. Parr and R. G. Pearson, *J. Am. Chem. Soc.*, 1983, **105**, 7512.
34 P. W. Ayers, R. G. Parr and R. G. Pearson, *J. Chem. Phys.*, 2006, **124**, 194107; R. G. Pearson, *Inorg. Chim. Acta*, 1995, **240**, 93.
35 P. Geerlings, F. De Proft and W. Langenaeker, *Chem. Rev.*, 2003, **103**, 1793.
36 T. K. Ghanty and S. K. Ghosh, *J. Phys. Chem.*, 1993, **97**, 4951; R. Roy, A. K. Chandra and S. Pal, *J. Phys. Chem.*, 1994, **98**, 10447; S. Hati and D. Datta, *J. Phys. Chem.*, 1994, **98**, 10451; Y. Simón-Manso and P. Fuenteaelba, *J. Phys. Chem. A*, 1998, **102**, 2029.
37 P. K. Chattaraj, U. Sarkar and D. R. Roy, *Chem. Rev.*, 2006, **106**, 2065.
38 R. G. Pearson, *J. Am. Chem. Soc.*, 1963, **85**, 3533; R. G. Pearson, *Hard and Soft Acids and Bases*, Dowen, Hutchinson and Ross, Stroudsberg, 1973; R. G. Pearson, *Chemical Hardness*, Wiley-VCH, New York, 1997.
39 R. G. Parr, L. V. Szentpály and S. Liu, *J. Am. Chem. Soc.*, 1999, **121**, 1922.
40 Computational Chemistry Comparison and Benchmark DataBase, National Institute of Standards and Technology http://cccbdb.nist.gov/.
41 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, 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, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Cli®ord, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *GAUSSIAN 03 (Revision D.01)*, Gaussian, Inc., Wallingford, CT, 2004.
42 J. M. Pérez-Jordá and E. San-Fabián, *Comput. Phys. Commun.*, 1993, **77**, 46; J. M. Pérez-Jordá, A. D. Becke and E. San-Fabián, *J. Chem. Phys.*, 1994, **100**, 6520.
43 F. Kurzer, *J. Chem. Educ.*, 2000, **77**, 851.

