

## INFORMATION ENERGY AND INFORMATION TEMPERATURE FOR MOLECULAR SYSTEMS

**Costinel Lepădatu and Emilia Nitulescu**

*Faculty of Pharmacy, University of Medicine and Pharmacy, Boulevard Maresal I. Antonescu, 94, Craiova, Romania*

*Received 15-10-2002*

### Abstract

The information energy and information temperature are proposed as new quantum descriptors for the characterization of molecular systems. Ab initio and semiempirical (PM3) procedures are used to create molecular probability fields on which the aforementioned information quantities are applied. The connection with other structural descriptors is discussed in the case of linear condensed aromatic rings and linear hydrocarbons with alternate double bonds.

### Introduction

The first mathematical function capable to describe the information content of a complex system has been proposed by Shannon related to the information content transmitted through different communication systems.<sup>1</sup> This quantity named information entropy is actually a convex function

$$f = -p \log_2 p,$$

defined onto the closed interval  $f : [0, 1] \rightarrow \mathbb{R}$ , for which Jensen's inequality<sup>2</sup>

$$\frac{1}{N} \sum_{n=1}^N f(p_n) \leq f\left(\frac{1}{N} \sum_{n=1}^N p_n\right), p_n \in [0,1]$$

should be valid.

Shannon proved that the function

$$S = -\sum_{n=1}^N p_n \log_2 p_n \tag{1}$$

named information entropy may characterize quantitatively very well the structure of a discrete set of "p<sub>n</sub>" values from the [0, 1] domain which may be considered as a probability field

$$\sum_{n=1}^N p_n = 1.$$

One may easily demonstrate that Shannon's entropy assumes a minimum value  $S_{\min} = 0$  for  $p_1 = 1$ ;  $p_2 = p_3 = \dots = 0$  and a maximum value for equal probabilities  $p_1 = p_2 = \dots = 1/N$ . These extreme two cases reflect the total order and total disorder respectively. This matter of facts justifies the name of entropy for the  $S$  function successfully applied to various domains, including chemistry and biology.<sup>2,3</sup>

The aim of this paper is to analyze two new information quantities capable to characterize the molecular systems, and namely, the information energy and temperature respectively.

Like the information entropy, the information energy introduced in mathematical statistics by Octav Onicescu<sup>4</sup> is a convex function defined onto  $[0, 1]$  interval, which may be a discrete or continuous probability field.

$$E = \sum_{n=1}^N p_n^2, p_n \in [0,1] \quad (2)$$

This function assumes for the two extreme cases total order and total disorder the maximum ( $E_{\max} = 1$ ) and minimum ( $E_{\min} = 1/N$ ) values respectively:

$$p_1 = 1, p_2 = p_3 = \dots = 0 \Rightarrow E_{\max} = 1 \text{ (total order)}$$

$$p_1 = p_2 = \dots = p_n = 1/N \Rightarrow E_{\min} = 1/N \text{ (total disorder).}$$

Because the function (2) reaches a minimum value for equal probabilities (total disorder), by analogy with thermodynamics, it has been called information energy. Also, Shannon's information entropy has to do with thermodynamic entropy. Jaynes has demonstrated the connection between Boltzmann entropy from statistical physics and information entropy.<sup>5</sup>

By analogy with thermodynamics, one may define the information temperature given by the ratio between Onicescu information energy and Shannon information entropy.

$$T = \frac{E}{S} \quad (3)$$

We shall try in the following, to analyze these three information quantities  $S$ ,  $E$  and  $T$  applied onto a probability field generated by the quantum molecular wave functions. The scope of this attempt is to see if these quantities may be used for the characterization of the molecular systems. Another goal is the possible connection between these quantities and another physico-chemical descriptors characterizing the molecular systems.

### Results and discussion

One knows from the quantum theory of the chemical bond (LCAO-MO approach) that the molecular wavefunctions (Molecular Orbitals)  $\psi_i$  are built from atomic orbitals  $\phi_j$

$$\psi_i = \sum_j c_{ij} \phi_j, \quad (4)$$

$c_{ij}$  being the mixing coefficients.

The probability field can be generated for each molecular electronic state, based on  $\int |\psi_i|^2 d\tau = 1$  and  $\int |\phi_j|^2 d\tau = 1$ . The mixing coefficients  $c_{ij}$  represent the measure of the atomic orbital contributions (weight) to the construction of the molecular orbitals. They are useful for the partition of electron population of the molecule usually made using two approximations: Mulliken's one where the electron densities from interatomic regions are equally shared to the two atoms participating to the chemical bonding (a kind of molecular democracy) and Löwdin's one which considers the baricenter of the interatomic electron density which is accordingly distributed to the two atoms. In Löwdin's approximation the more electronegative atom receives more interatomic electron density than the less electronegative atom (a kind of molecular liberalism).

The probability field characterizing the molecular system can be obtained in different ways depending on the partition of the electron density:

(1) For each atomic orbital  $(\phi_j)^{\rho_j}$ , the  $\rho_j$  electron density is allocated using either Mulliken or Löwdin procedure, comprised according to Pauli aufbau principle within  $0 \leq \rho_j \leq 2$ . In this case, the following probability field may be built up, as follows:

$p_j = \frac{\rho_j}{2}$  and  $\overline{p_j} = 1 - p_j$ , so that  $p_j + \overline{p_j} = 1$ , where  $\overline{p_j}$  means the probability that  $\rho_j$  electron density does not exist on  $\phi_j$  atomic orbital. We get in this way the orbital information entropy (S), energy (E) and temperature (T), if formulas 1–3 are applied onto such a probability field.

(2) The second possibility refers to the partition of all electrons in molecule on all atomic orbitals:

$$\sum_i \rho_i = N, \quad (5)$$

where  $\rho_i$  is the electron density on every atomic orbital, as results from the summation of all contributions of that atomic orbital to those molecular orbitals “occupied” with electrons. For ab initio procedure all atomic orbitals are considered. In this case N should represent the total number of electrons in molecule. For semiempirical approximations, only the valence atomic orbitals and valence electrons in molecule are considered.

From (5) we can generate a probability field  $p_i = \frac{\rho_i}{N}; \sum_i p_i = 1$ .

The S, E, T information functions (formulas 1–3) applied onto such a probability field will be called molecular entropy (S), energy (E) and temperature (T).

(3) The third possibility is to create for each atom a probability field

$$\left| \varphi_1^{\rho_1} \varphi_2^{\rho_2} \dots \varphi_i^{\rho_i} \right|_{atom}; \rho_1 + \rho_2 + \dots + \rho_i = \rho_j (atom),$$

where  $\rho_1, \rho_2, \dots, \rho_i$  are electron densities distributed on the atomic orbital of “j” atom,  $\rho_j$  represents the electron density of that atom, so that  $\sum_j \rho_j = N$ , where N is the number

of electrons in molecule (ab initio) or the number of valence electrons (semiempirical approximations). One may define a probability field for each “j” atom in molecule,

$p_j = \frac{\rho_j}{N}; \overline{p_j} = 1 - p_j; p_j + \overline{p_j} = 1$ , where  $\overline{p_j}$  means the probability that  $\rho_j$  electron

density does not exist on “j” atom. The S, E, T functions applied onto such a probability field will be called atomic information entropy (S), energy (E) and temperature (T).

Let us discuss in the following, the first two possibilities, and namely the molecular and orbital information S, E, T functions as applied to a series of linear with alternate double bonds hydrocarbons and to a series of condensed aromatic hydrocarbons. The quantum molecular calculations were performed by means of GAMESS (ab initio) and MOPAC 7 (semiempirical) offered as shareware programs for academic researches.<sup>6,7</sup>

The molecular energy levels, electron population distributed on AO's, MO's have been calculated, as well as dipole moments, polarizabilities etc. have been calculated by means of these programs. The analysis of the results and the calculation of information entropy, energy and temperature have been done using specially dedicated programs developed by us.

As may be seen in Table 1, the values for the molecular information entropy continuously decreases with the increase of the number of carbon atoms. In exchange, the orbital temperature increases separately in the sets of hydrocarbons with odd and even number of carbon atoms. It is worth to note that the sums of orbital entropies and energies notably increase with the increase of carbon atoms of the hydrocarbons. In exchange, the corresponding orbital temperature values remain almost constant. The correlation between the information quantities S, E and T and other quantum molecular descriptors characterizing the molecular structures have also been performed. Such a correlation has been done in order to see the possible connection between the information quantities herein defined and other physical and chemical quantities representing the molecules.

**Table 1.** Information quantities S, E and T calculated from ab initio results (GAMESS, RHF, MP4, STO-6G, Löwdin) for linear hydrocarbons.

Molecule	S	E	T	$\Sigma S$	$\Sigma E$	$T = \Sigma E / \Sigma S$
	(molecular)			(orbital)		
C <sub>2</sub> H <sub>4</sub>	3.75085	0.07800	0.02080	12.08473	7.98449	0.66071
C <sub>3</sub> H <sub>6</sub>	4.33590	0.05200	0.01199	18.13502	11.99466	0.66031
C <sub>4</sub> H <sub>6</sub>	4.64125	0.04215	0.00908	22.17254	14.96649	0.67500
C <sub>5</sub> H <sub>8</sub>	4.98579	0.03318	0.00666	28.22144	18.95725	0.67173
C <sub>6</sub> H <sub>8</sub>	5.18777	0.02887	0.00557	32.25860	21.94943	0.68042
C <sub>7</sub> H <sub>10</sub>	5.43228	0.02436	0.00448	38.30763	25.93976	0.67716
C <sub>8</sub> H <sub>10</sub>	5.58323	0.02196	0.00393	42.34547	28.93134	0.68322
C <sub>9</sub> H <sub>12</sub>	5.77273	0.01925	0.00333	48.40075	32.92108	0.68018
C <sub>10</sub> H <sub>12</sub>	5.89330	0.01771	0.00301	52.44267	35.9140	0.68477
C <sub>20</sub> H <sub>22</sub>	6.86921	0.00901	0.00131	102.8644	70.82584	0.68854

$\Sigma S$ ,  $\Sigma E$  = sum of orbital information entropies or energies.

As may be seen in Table 2, the linear regression between various quantum molecular descriptors and information quantities S, E, T shows that the molecular information temperature correlates well with the average electrophilic reaction index for the carbon atoms in the series of studied hydrocarbons ( $R_2 = 0.9954$ ) and fairly good with the ionization potentials which are the highest occupied molecular orbital (HOMO) energy level ( $R_2 = 0.8777$ ).

**Table 2.** Linear correlation  $y=a_0+b_1x$  between the information quantities S, E, T and quantum molecular descriptors for linear hydrocarbons.

y	x	R <sub>2</sub>
T (molecular)	Average electrophilic reaction index E <sub>C</sub> for the carbon atoms	0.9954
	HOMO energy levels	0.8777
T (orbital)	Average electrophilic reaction index E <sub>C</sub> for the carbon atoms(only for hydrocarbons with odd number of carbon atoms)	0.9999
	Average nucleophilic reaction index N <sub>C</sub> for the carbon atoms(only for hydrocarbons with even number of carbon atoms)	0.9999
E (molecular)	Average electrophilic reaction index E <sub>C</sub> for the carbon atoms	0.9994
	LUMO energy levels	0.9188
ΣE (orbital)	HOMO – 1 energy levels	0.9568

R<sub>2</sub> = correlation coefficient.

The orbital temperature correlates very well with the average electrophilic reaction index of the carbon atoms in the hydrocarbons with odd number of carbon atoms and with average nucleophilic reaction index N<sub>C</sub> in the hydrocarbons with even number of carbon atoms.

The reactivity indices are defined as<sup>8</sup>

$$E_A = \sum_{i \in A} c_{i_{HOMO}}^2 \quad (\text{electrophilic})$$

and

$$N_A = \sum_{j \in A} c_{j_{LUMO}}^2 \quad (\text{nucleophilic}),$$

where the summation are done over all AO's of a given atom species,  $c_{i_{HOMO}}$  and  $c_{j_{LUMO}}$  being the AO coefficients on the HOMO and LUMO respectively. These reactivity indices are related to the activation energy of the corresponding chemical reaction (electrophilic and nucleophilic).

Onicescu molecular information energy correlates with the electrophilic reaction index E<sub>C</sub> of the carbon atoms for all linear hydrocarbons (R<sub>2</sub> = 0.9994) and also with affinity (LUMO energy level) (R<sub>2</sub> = 0.9188). The sum of orbital information energies correlate with the HOMO-1 energy levels which are the second ionization potential of the molecule (R<sub>2</sub> = 0.9568).

It is interesting to note that the sum of orbital information energies values ΣE(orbital) are close to the number of occupied energy levels in the molecule. As an example, for C<sub>7</sub>H<sub>10</sub> every carbon atom participates with 6 electrons and every hydrogen

atom with one electron in ab initio interactions, the number of occupied molecular energy levels is equal to 26. In this case  $\Sigma E = 25.939$  (see Table 1).

For the second series of linearly condensed aromatic hydrocarbons (benzene, naphthalene, anthracene etc.) the semiempirical results are summarized in Table 3.

**Table 3.** Information quantities S, E and T calculated from semiempirical results MOPAC 7 (RHF, MNDO, Mulliken) for condensed aromatic hydrocarbons.

Molecule	S	E (molecular)	T=E/S	$\Sigma S$	$\Sigma E$ (orbital)	T= $\Sigma E/\Sigma S$
CY1	4.89868	0.03373	0.00689	29.74158	15.17808	0.51033
CY2	5.57657	0.02109	0.00378	47.57882	24.29032	0.51053
CY3	6.03592	0.01534	0.00254	65.41570	33.40282	0.51062
CY4	6.38379	0.01205	0.00189	83.25190	42.51576	0.51069
CY5	6.66386	0.00992	0.00149	101.0886	51.62837	0.51072
CY6	6.89830	0.00844	0.00122	118.9243	60.74168	0.51076

CY1 = benzene; CY2 = naphthalene; CY3 = anthracene etc.

As may be seen in Table 3, the studied information quantities vary almost monotonously irrespective of the parity of the condensed rings number. As in the case of linear hydrocarbons with alternate double bonds, the sum of orbital information energies  $\Sigma E(\text{orbital})$  approximates fairly well the number of occupied molecular levels.

As an example, for the molecule with six linearly condensed aromatic rings (CY6), the total number of valence electrons is equal to 120 and hence the number of occupied molecular levels 60, comparable with  $\Sigma E = 60.74$ . This observation suggests a way to approximate  $\Sigma E(\text{orbital})$  for both ab initio and PM3 procedures with the number of occupied molecular levels in the molecule. The linear correlation of the analyzed information quantities S, E, T with different quantum molecular descriptors is given in Table 4.

**Table 4.** Linear correlation  $y=a_0+b_1x$  between the information quantities S, E, T and quantum molecular descriptors for condensed aromatic hydrocarbons.

y	x	$R_2$
T (molecular)	LUMO + 1	0.9271
T (orbital)	FPSA	0.9779
E (molecular)	Average electrophilic reaction index for C atoms	0.9999
	LUMO + 1	0.9487
$\Sigma E$ (orbital)	HOMO - 1	0.9627

$R_2$  = correlation coefficient.

As may be seen in Table 4, the molecular information energy correlates very well with the average electrophilic reaction index for the carbon atoms in molecule ( $R_2 = 0.999$ ). The same situation has been observed with the linear hydrocarbons (Table 4).

The molecular information energy as well as the molecular information temperature correlate fairly good with LUMO+1 energy level, which is the second affinity energy of the molecule ( $R_2 = 0.9627$  and  $0.9271$  respectively). The orbital information temperature correlates very well ( $R_2 = 0.9779$ ) with an important molecular descriptor FPSA equal to the fraction between the Partial Positive Surface Area of the molecule (PPSA) and the Total Molecular Surface Area (TMSA). This descriptor belongs to the charged partial surface area descriptors<sup>9,10</sup> may describe the polar interactions between molecules or receptor–ligand interactions for biomolecules or drugs.

This kind of researches regarding the information entropy, energy and temperature and their correlations with other physical and chemical descriptors will be extended to other classes of substances.

### References

1. C. Shannon, W. Weaver, *Mathematical Theory of Communication*, Illinois University, Illinois Press, Urbana, **1949**.
2. N. Rashevsky, *Bull. Math. Biophysics* **1955**, *17*, 229–235.
3. G. Karreman, *Bull. Math. Biophysics* **1955**, *17*, 279–285.
4. O. Onicescu V. Stefanescu, *Elemente de statistica informatională cu aplicatii*, Ed. Tehnica, Bucharest, **1979**.
5. E. T. Jaynes, *Phys. Rev.* **1957**, *106*, 620–630.
6. M. W. Schmidt, K. K. Baldrige, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis, J. A. Montgomery, *J. Comput. Chem.* **1993**, *14*, 1347–1363.
7. J. J. P. Stewart, *J. Comp. Chem.* **1989**, *10*, 209–220.
8. K. Fukui, *Theory of Orientation and Stereoselection*, Springer Verlag, Berlin, **1975**.
9. D. T. Stanton, P.C.Jurs, *Anal. Chem.*; **1990**, *62*, 2323–2329.
10. D. T. Stanton, L. M. Egalf, P. C. Jurs, M. G. Hicks, *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 306–316.

### Povzetek

V delu predlagamo informacijsko energijo in informacijsko temperaturo kot nova kvantnokemijska deskriptorja za karakterizacijo molekularnih sistemov. Ab initio in semiempirično (PM3) metodologijo smo uporabili za izračun molekularnih verjetnostnih polj, na katerih sta bili določeni omenjeni količini. Povezanost z drugimi strukturnimi deskriptorji je v delu obravnavana na primeru linearno kondenziranih aromatskih obročev in linearnih ogljikovodikov z alternirajočimi dvojnimi vezmi.