Decoding the Building Blocks of Life from the Perspective of Quantum Information

Rodolfo O. Esquivel, Moyocoyani Molina-Espíritu,
Frank Salas, Catalina Soriano, Carolina Barrientos,
Jesús S. Dehesa and José A. Dobado

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/55160

1. Introduction

Physical theories often start out as theories which only embrace essential features of the macroscopic world, where their predictions depend on certain parameters that have to be either assumed or taken from experiments; as a result these parameters cannot be predicted by such theories. To understand why the parameters have the values they do, we have to go one level deeper—typically to smaller scales where the easiest processes to study are the ones at the lowest level. When the deeper level reduces the number of unknown parameters, we consider the theory to be complete and satisfactory. The level below conventional molecular biology is spanned by atomic and molecular structure and by quantum dynamics. However, it is also true that at the lowest level it becomes very difficult to grasp all the features of the molecular processes that occur in living systems such that the complexity of the numerous parameters that are involved make the endeavour a very intricate one. Information theory provides a powerful framework for extracting essential features of complicated processes of life, and then analyzing them in a systematic manner. In connection to the latter, quantum information biology is a new field of scientific inquiry in which information-theoretical tools and concepts are permitting to get insight into some of the most basic and yet unsolved questions of molecular biology.

Chirality is often glossed over in theoretical or experimental discussions concerning the origin of life, but the ubiquity of homochiral building blocks in known biological systems demands explanation. Information theory can provide a quantitative framework for understanding the role of chirality in biology. So far it has been thought that the genetic code is "unknowable"

by considering DNA as a string of letters only (... ATTGCAAGC...) and likewise by considering proteins as strings of identifiers (... DYRFQ...), we believe that this particular conclusion might be probably wrong because it entirely fails to consider the information content of the molecular structures themselves and their conformations.

On the other hand, according to molecular biology, living systems consist of building blocks which are encoded in nucleic acids (DNA and RNA) and proteins, which possess complex patterns that control all biological functions. Despite the fact that natural processes select particular building blocks which possess chemical simplicity (for easy availability and quick synthesis) and functional ability (for implementing the desired tasks), the most intriguing question resides in the amino acid selectivity towards a specific codon/anticodon. The universal triplet genetic code has considerable and non-uniform degeneracy, with 64 codons carrying 21 signals (including Stop) as shown in Table 1. Although there is a rough rule of similar codons for similar amino acids, no clear pattern is obvious.

Information theory of quantum many-body systems is at the borderline of the development of physical sciences, in which major areas of research are interconnected, i.e., physics, mathematics, chemistry, and biology. Therefore, there is an inherent interest for applying theoretic-information ideas and methodologies to chemical, mesoscopic and biological systems along with the processes they exert. On the other hand, in recent years there has been an increasing interest in applying complexity concepts to study physical, 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 Kolmogorov–Chaitin or algorithmic information theory (Kolmogorov, 1965; Chaitin, 1966), the classical information theory of Shannon and Weaver (Shannon & Weaver, 1948), Fisher information (Fisher, 1925; Frieden, 2004), and the logical (Bennet, 1988) and the thermodynamical (Lloyd & Pagels, 1988) depths, among others. Some of them share rigorous connections with others as well as with Bayes and information theory (Vitanyi & Li, 2000). The term complexity has been applied with different meanings: algorithmic, geometrical, computational, stochastic, effective, statistical, and structural among others and it has been employed in many fields: dynamical systems, disordered systems, spatial patterns, language, multielectronic systems, cellular automata, neuronal networks, self-organization, DNA analyses, social sciences, among others (Shalizi et al., 2004; Rosso et al., 2003; Chatzisavvas et al., 2005; Borgoo et al., 2007).

The definition of complexity is not unique, its quantitative characterization has been an important subject of research and it has received considerable attention (Feldman & Crutchfield, 1998; Lamberti et al., 2004). The usefulness of each definition depends on the type of system or process under study, the level of the description, and the scale of the interactions among either 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 like 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 might intervene in the definitions so as to quantita-

tively 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 delocalization/uncertainty. This is the case of the definition of López-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 (Anteonodo & Plastino, 1996), modified (Catalán et al., 2002; Martin et al., 2003) and generalized (López-Ruiz, 2005) leading to a useful estimator which satisfies several desirable properties of invariance under scaling transfromations, translation, and replication (Yamano, 2004; Yamano, 1995). The utility of this improved complexity has been verified in many fields [8] and allows reliable detection of periodic, quasiperiodic, linear stochastic, and chaotic dynamics (Yamano, 2004; López-Ruiz et al., 1995; Yamano, 1995). 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 (Carbó-Dorca et al., 1980) or information energy Onicescu, 1996), which quantifies the departure of the probability density from uniformity (Catalán et al., 2002; Martin et al., 2003) (equiprobability) and the Shannon entropy S, which is a general measure of randomness/uncertainty of the probability density (Shannon & Weaver, 1948), and quantifies the departure of the probability density from localizability. Both global quantities are closely related to the measure of spread of a probability distribution.

The Fisher-Shannon product FS has been employed as a measure of atomic correlation (Romera & Dehesa, 2004) and also defined as a statistical complexity measure (Angulo et al., 2008a; Sen et al., 2007a). The product of the power entropy J -explicitly defined in terms of the Shannon entropy (see below)- and the Fisher information measure, I, combine both the global character (depending on the distribution as a whole) and the local one (in terms of the gradient of the distribution), to preserve the general complexity properties. As compared to the LMC complexity, aside of the explicit dependence on the Shannon entropy which serves to measure the uncertainty (localizability) of the distribution, the Fisher-Shannon complexity replaces the disequilibrium global factor D by the Fisher local one to quantify the departure of the probability density from disorder (Fisher, 1925; Frieden, 2004) of a given system through the gradient of the distribution.

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 (Fisher, 1925; Frieden, 2004), as well as the time-independent Kohn-Sham equations and the time-dependent Euler equation (Nagy, 2003; Nalewajski, 2003). More recently, the Fisher information has been employed also as an intrinsic accuracy measure for specific atomic models and densities (Nagy & Sen, 2006; Sen et al., 2007b)), as well as for general quantum-mechanical central potentials (Romera et al. 2006; Dehesa et al., 2007). The concept of phase-space Fisher information has been analyzed for hydrogenlike atoms and the isotropic harmonic oscillator (Hornyak & Nagy, 2007), where both position and momentum variables are included. Several applications concern atomic distributions in position and momentum spaces have been performed where the FS complexity is shown to provide relevant information on atomic shell structure and ionization processes (Angulo et al., 2008a; Sen et al., 2007a; Angulo & Antolín, 2008b; Antolín & Angulo, 2009).

In line with the aforementioned developments we have undertaken multidisciplinary research projects so as to employ IT at different levels, classical (Shannon, Fisher, complexity, etc) and quantum (von Neumann and other entanglement measures) on a variety of chemical process-es, organic and nanostructured molecules. Recently, significant advances in chemistry have been achieved by use of Shannon entropies through the localized/delocalized features of the electron distributions allowing a phenomenological description of the course of elementary chemical reactions by revealing important chemical regions that are not present in the energy profile such as the ones in which bond forming and bond breaking occur (Esquivel et al., 2009). Further, the synchronous reaction mechanism of a S_N2 type chemical reaction and the non-synchronous mechanistic behavior of the simplest hydrogenic abstraction reaction were predicted by use of Shannon entropies analysis (Esquivel et al., 2010a). In addition, a recent study on the three-center insertion reaction of silylene has shown that the information-theoretical measures provide evidence to support the concept of a continuum of transient of Zewail and Polanyi for the transition state rather than a single state, which is also in agreement with other analyses (Esquivel et al., 2010b). While the Shannon entropy has remained the major tool in IT, there have been numerous applications of Fisher information through the "narrow-ness/disorder" features of electron densities in conjugated spaces. Thus, in chemical reactions the Fisher measure has been employed to analyze its local features (Esquivel et al., 2010c) and also to study the steric effect of the conformational barrier of ethane (Esquivel et al., 2011a). Complexity of the physical, chemical and biological systems is a topic of great contemporary interest. The quantification of complexity of real systems is a formidable task, although various single and composite information-theoretic measures have been proposed. For instance, Shannon entropy (S) and the Fisher information measure (I) of the probability distributions are becoming increasingly important tools of scientific analysis in a variety of disciplines. Overall, these studies suggest that both S and I can be used as complementary tools to describe the information behavior, pattern, or complexity of physical and chemical systems and the electronic processes involving them. Besides, the disequilibrium (D), defined as the expectation value of the probability density is yet another complementary tool to study complexity since it measures its departure from equiprobability. Thus, measuring the complexity of atoms and molecules represents an interesting area of contemporary research which has roots in infor-mation theory (Angulo et al., 2010d). In particular, complexity measures defined as products of S and D or S and I have proven useful to analyze complexity features such as order, uncertainty and pattern of molecular systems (Esquivel et al., 2010f) and chemical processes (Esquivel et al., 2011b). On the other hand, the most interesting technological implications of quantum mechanics are based on the notion of entanglement, which is the essential ingredient for the technological implementations that are foreseen in the XXI century. Up to now it remains an open question whether entanglement can be realized with molecules or not and hence it is evident that the new quantum techniques enter the sphere of chemical interest. Generally speaking, entanglement shows up in cases where a former unit dissociates into simpler sub-systems, the corresponding processes are known quite well in chemistry. Al-though information entropies have been employed in quantum chemistry, applications of entanglement measures in chemical systems are very scarce. Recently, von Neumann measures in Hilbert space have been proposed and applied to small chemical systems (Carrera et al.

2010, Flores-Gallegos and Esquivel, 2008), showing than entanglement can be realized in molecules. For nanostructures, we have been able to show that IT measures can be successfully employed to analyse the growing behaviour of PAMAM dendrimers supporting the dense-core model against the hollow-core one (Esquivel et al., 2009b, 2010g, 2011c).

In the Chapter we will present arguments based on the information content of L- and D-aminoacids to explain the biological preference toward homochirality. Besides, we present benchmark results for the information content of codons and aminoacids based on information-theoretical measures and statistical complexity factors which allow to elucidate the coding links between these building blocks and their selectivity.

2. Information-theoretical measures and complexities

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, (p), is obtained through the molecular momentals (momentum-space orbitals) $\phi_i(p)$, and similarly for the position-space density, $\rho(r)$, through the molecular position-space orbitals $\phi_i(r)$. The momentals can be obtained by three-dimensional Fourier transformation of the corresponding orbitals (and conversely)

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

Standard procedures for the Fourier transformation of position space orbitals generated by ab-initio methods have been described (Rawlings & Davidson, 1985). 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 (Kaijser & Smith, 1997), the transformation of the total molecular electronic wavefunction from position to momentum space is computationally straightforward (Kohout, 2007).

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, $\rho(r)$, the C(*LMC*) complexity is given by (Feldman & Crutchfield, 1998; Lamberti et al., 2004; Anteonodo & Plastino, 1996; Catalán et al., 2002; Martin et al., 2003):

$$C_r(LMC) = D_r e^{S_r} = D_r L_r \tag{2}$$

where *Dr* is the disequilibrium (Carbó-Dorca et al., 1980; Onicescu, 1966)

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

and S is the Shannon entropy (Shannon & Weaver, 1949)

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

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

It is important to mention that the *LMC* complexity of a system must comply with the following lower bound (López-Rosa et al., 2009):

$$C_p(LMC) = D_p e^{Sp} = D_p L_p \tag{5}$$

The *FS* complexity in position space, *Cr(FS)*, is defined in terms of the product of the Fisher information (Fisher, 1925; Frieden, 2004)

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

and the power entropy (Angulo et al. 2008a; Sen et al., 2007a) in position space, *Jr*

$$I_r = \int \rho(\mathbf{r})\left|\bar{\nabla}\ln\rho(\mathbf{r})\right|^2 d^3\mathbf{r} \tag{7}$$

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

$$J_r = \frac{1}{2\pi e}e^{\frac{2}{3}S_r}, \tag{8}$$

and similarly

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

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 relationship (Dembo et al., 1991).

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

with n being the space dimensionality, thus providing a universal lower bound to FS complexity. The definition in Eq. (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 3-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 rp-space, characterized by the probability density $f(r, p) = \rho(r)\gamma(p)$, where the complexity measures are defined as

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

and

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

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 the momentum space contribution plays a more relevant role as compared to the one in position space.

We have also evaluated some reactivity parameters that may be useful to analyze the chemical reactivity of the aminoacids. So that, we have computed several reactivity properties such as the ionization potential (IP), the hardness (η) and the electrophilicity index (ω). These properties were obtained at the Hartree-Fock level of theory (HF) in order to employ the Koopmans' theorem (Koopmans, 1933; Janak, 1978), 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. Parr and Pearson, proposed a quantitative definition of hardness (η) within conceptual DFT (Parr & Yang, 1989):

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

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 Eq. (14) was put originally to make the hardness definition symmetrical with respect to the chemical potential (Parr & Pearson, 1983)

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

although it has been recently disowned (Ayer et al. 2006: Pearson, 1995). In general terms, the chemical hardness and softness are good descriptors of chemical reactivity. The former has been employed (Ayer et al. 2006: Pearson, 1995; Geerlings et al., 2003) 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 (Ghanty & Ghosh, 1993; Roy et al., 1994; Hati & Datta, 1994; Simon-Manso & Fuentealba, 1998) and hence soft molecules are more polarizable and possess predisposition to acquire additional electronic charge (Chattaraj et al., 2006). The chemical hardness η is a central quantity for use in the study of reactivity through the hard and soft acids and bases principle (Pearson, 1963; Pearson, 1973; Pearson, 1997).

The electrophilicity index (Parr et al., 1999), 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

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

The electrophilicity is also a good descriptor of chemical reactivity, which quantifies the global electrophilic power of the molecules -predisposition to acquire an additional electronic charge- (Parr & Yang, 1989).

3. Aminoacids

The exact origin of homochirality is one of the great unanswered questions in evolutionary science; such that, the homochirality in molecules has remained as a mystery for many years ago, since Pasteur. Any biological system is mostly composed of homochiral molecules; therefore, the most well-known examples of homochirality is the fact that natural proteins are composed of L-amino acids, whereas nucleic acids (RNA or DNA) are composed of D-sugars (Root-Bernstein, 2007; Werner, 2009; Viedma et al., 2008). The reason for this behavior continues to be a mystery. Until today not satisfactory explanations have been provided regarding the origin of the homochirality of biological systems; since, the homochirality of the amino acids is critical to their function in the proteins. If proteins (with L-aminoacids) had a

non-homochiral behavior (with few D-enantiomers in random positions) they would not present biological functionality It is interesting to mention that L-aminoacids can be synthesized by use of specific enzymes, however, in prebiotic life these processes remain unknown. The same problem exists for sugars which have the D configuration. (Hein and Blackmond, 2011; Zehnacker et al., 2008; Nanda and DeGrado, 2004).

On the other hand, the natural amino acids contain one or more asymmetric carbon atoms, except the glycine. Therefore, the molecules are two nonsuperposable mirror images of each other; i.e., representing right-handed (D enantiomer) and left-handed (L enantiomer) structures. It is considered that the equal amounts of D- and L- amino acids existed on primal earth before the emergence of life. Although the chemical and physical properties of L-and D amino acids are extremely similar except for their optical character, the reason of the exclusion of D-amino acids and why all living organisms are now composed predominantly of L-amino acids are not well-known: however, the homochirality is essential for the development and maintenance of life (Breslow, 2011; Fujii et al., 2010; Tamura, 2008). The essential property of α-aminoacids is to form linear polymers capable of folding into 3-dimensional structures, which form catalytic active sites that are essential for life. In the procees, aminoacids behave as hetero bifunctional molecules, forming polymers via head to tail linkage. In contrast, industrial nylons are often prepared from pairs of homo-bifunctional molecules (such as diamines and dicarboxylic acids), the use of a single molecule containing both linkable functionalities is somewhat simpler (Cleaves, 2010; Weber and Miller, 1981; Hicks, 2002).

The concept of chirality in chemistry is of paramount interest because living systems are formed of chiral molecules of biochemistry is chiral (Proteins, DNA, amino acids, sugars and many natural products such as steroids, hormones, and pheromones possess chirality). Indeed, amino acids are largely found to be homochiral (Stryer, 1995) in the L form. On the other hand, most biological receptors and membranes are chiral, many drugs, herbicides, pesticides and other biological agents must themselves possess chirality. Synthetic processes ordinarily produce a 50:50 (racemic) mixture of left-handed and right-handed molecules (so-called enantiomers), and often the two enantiomers behave differently in a biological system.

On the other hand, a major topic of research has been to study the origin of homochirality. In this respect, biomembranes have played an important role for the homochiraility of biopolymers. One of the most intriguing problems in life sciences is the mechanism of symmetry breaking. Many theories have been proposed on these topics and in the attempt to explain the amplification of a first enantiomeric imbalance to the enantiopurity of biomolecules (Bombelli et al., 2004). In all theories on symmetry breaking and on enantiomeric excess amplification little attention has been paid to the possible role of biomembranes, or of simple self-aggregated systems that may have acted as primitive biomembranes. Nevertheless, it is possible that amphiphilic boundary systems, which are considered by many scientists as intimately connected to the emergence and the development of life (Avalos et al. 2000; Bachmann et al., 1992), had played a role in the history of homochirality in virtue of recognition and compartmentalization phenomena (Menger and Angelova, 1998). In general, the major reason for the different recognition of two enantiomers by biological cells is the homochirality of biomolecules such as L-amino acids and D-sugars. The diastereomeric interaction between the

enantiomers of a bioactive compound and the receptor formed from a chiral protein can cause different physiological responses. The production technology of enantiomerically enriched bioactive compounds one of the most important topics in chemistry. There is great interest in how and when biomolecules achieved high enantioenrichment, including the origin of chirality from the standpoint of chiral chemistry (Zehnacker et al., 2008; Breslow, 2011; Fujii et al., 2010; Tamura, 2008; Arnett and Thompson, 1981)

3.1. Physical and information-theoretical properties

Figure l illustrates a Venn diagram (Livingstone & Barton, 1993; Betts & Russell, 2003) which is contained within a boundary that symbolizes the universal set of 20 common amino acids (in one letter code). The amino acids that possess the dominant properties—hydrophobic, polar and small (< 60 Å3)—are defined by their set boundaries. Subsets contain amino acids with the properties aliphatic (branched sidechain non-polar), aromatic, charged, positive, negative and tiny (<35 Å 3). Shaded areas define sets of properties possessed by none of the common amino acids. For instance, cysteine occurs at two different positions in the Venn diagram. When participating in a disulphide bridge (C_{S-S}), cysteine exhibits the properties 'hydrophobic' and 'small'. In addition to these properties, the reduced form (C_{S-H}) shows polar character and fits the criteria for membership of the 'tiny' set. Hence, the Venn diagram (Figure l) assigns multiple properties to each amino acid; thus lysine has the property hydrophobic by virtue of its long sidechain as well as the properties polar, positive and charged. Alternative property tables may also be defined. For example, the amino acids might simply be grouped into non-intersecting sets labelled, hydrophobic, charged and neutral.
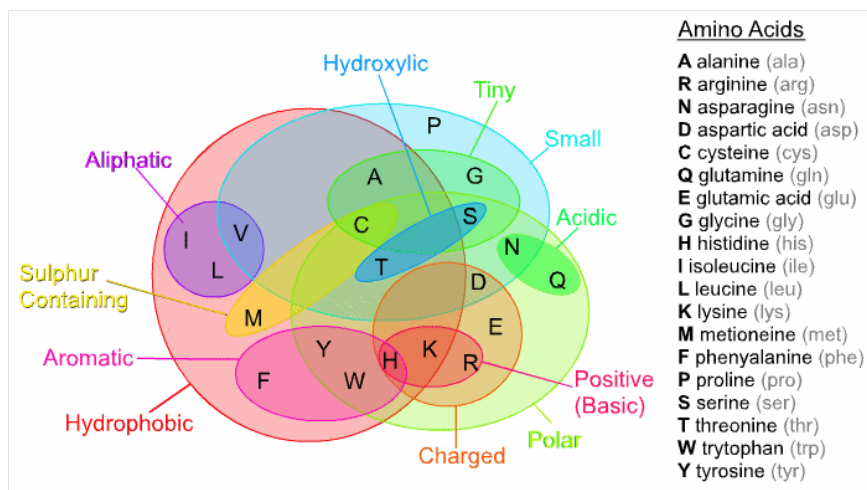
Figure 1. Venn diagram (Livingstone & Barton, 1993; Betts & Russell, 2003) of boundaries that symbolizes the universal set of 20 common amino acids (b). The Venn diagram in Fig. (1a) may be simply encoded as the property table or index shown here where the rows define properties and the columns refer to each amino acid.

In order to perform a theoretical-information analysis of L- and D-aminoacids we have employed the corresponfing L-enantiomers reported in the Protein Data Bank (PDB), which provide a standard representation for macromolecular structure data derived from X-ray diffraction and NMR studies. In a second stage, the D-type enantiomers were obtained from the L-aminoacids by interchanging the corresponding functional groups (carboxyl and amino) of the α-carbon so as to represent the D-configuration of the chiral center, provided that steric impediments are taken into account. The latter is achieved by employing the Ramachandran (Ramachandran et al, 1963) map, which represent the phi-psi torsion angles for all residues in the aminoacid structure to avoid the steric hindrance. Hence, the backbone of all of the studied aminoacids represent possible biological structures within the allowed regions of the Rama-chandran. In the third stage, an electronic structure optimization of the geometry was per-formed on all the enantiomers for the twenty essential aminoacids so as to obtain structures of minimum energy which preserve the backbone (see above). In the last stage, all of the information-theoretic measures were calculated by use of a suite of programs which have been discussed elsewhere (Esquivel et al., 2012).

In Figures 2 through 4 we have depicted some selected information-theoretical measures and complexities in position space versus the number of electrons and the energy. For instance, it might be observed from Fig. 2 that the Shannon entropy increases with the number of electrons so that interesting properties can be observed, e.g., the aromatic ones possess more delocalized densities as the rest of the aminoacids (see Figure 1B) which confer specific chemical properties. On the other hand, the disequilibrium diminishes as the number of electron increases (see Fig. 2), which can be related to the chemical stability of the aminoacids, e.g., cysteine and metionine show the larger values (see Fig. 2) which is in agreement with the biological evidence in that both molecules play mutiple functions in proteins, chemical as well as structural, conferring the higher reactivity that is recognized to both molecules. In contrast, aromatic aminoacids (see Fig 1B) are the least reactive, which is in agreement with the lower disequilibrium values that are observed form Fig 2.
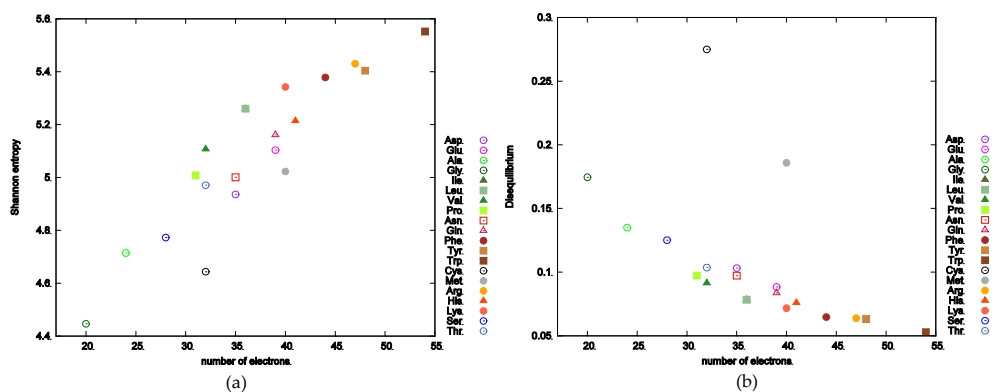
Figure 2. Shannon entropy (left) and disequilibrum (right), both in position space, versus the number of electrons, for the set of 20 aminoacids.
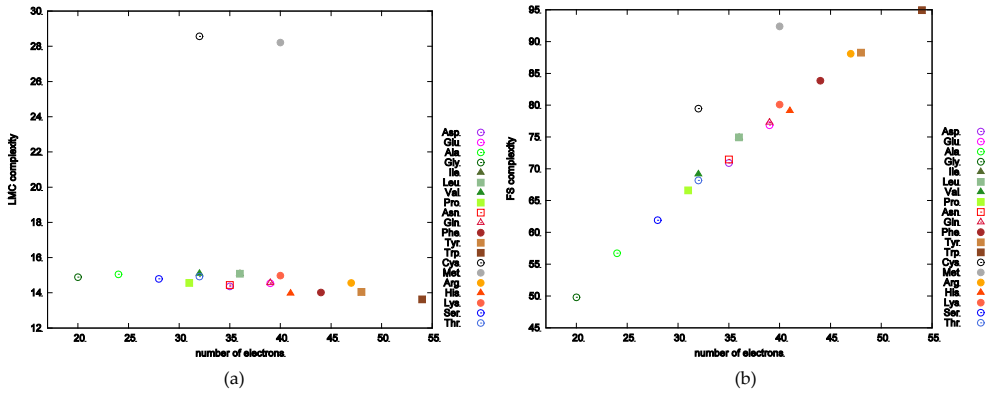
Figure 3. LMC- (left) and FS-complexities (right), both in position space, versus the number of electrons, for the set of 20 aminoacids.
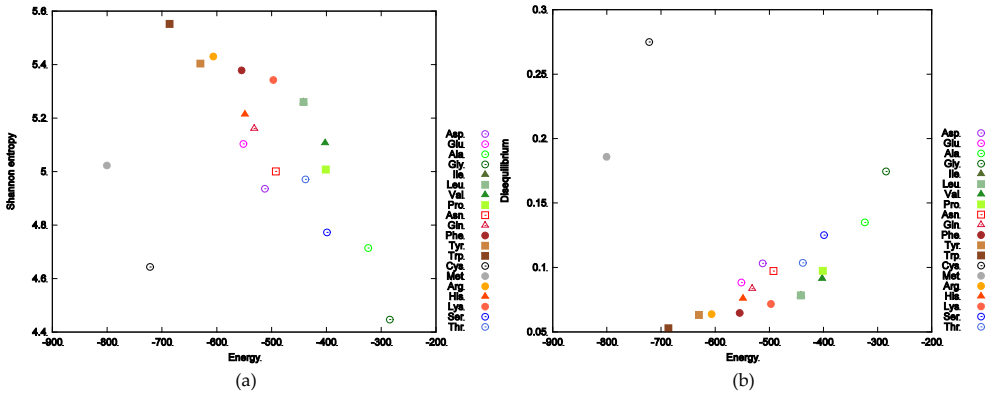

Figure 4. LMC- (left) and FS-complexities (right), both in position space, versus the total energy, for the set of 20 aminoacids.

In Figures 3 we have plotted the *LMC* and *FS* complexities versus the number of electrons for the twenty aminoacids where we can observe that *LMC* complexity disntinguishes two different groups of aminoacids, where the more reactive (met and cys) hold larger values. In contrast, *FS* complexity behaves linearly with the number of electrons where the aromatic aminoacids possess the larger values and hence represent the more complex ones. Furthemore,

the behavior of the *LMC* and *FS* complexities with respect to the total energy is analyzed in Figures 4, to note that *LMC* complexity characterizes two different groups of aminoacids where the most reactive (cys and met) possess the largest values, which incidentally hold the largest energies (negatively). A different behavior is observed for the *FS* complexity in that the smaller values correspond to the less energetic aminoacids. It is worthy to mention that the *FS* complexity is related to the Fisher information measure (Eq. 7) which depends on the local behavior of the position space density, i..e., simpler molecules present more ordered chemical structures, and hence these kind of aminoacids are expected to be less complex, e.g., the small and the tiny ones (Ser, Ala, Thr).

3.2. Homochirality

In Figures 5 through 8 we have analyzed the homochiral behavior of all aminoacids by plotting the difference between the L and the D values of several physical properties (energy, ionization potential, hardness, electrophilicity) and some relevant information-theoretical measures (Shannon entropy, Fisher, LMC- and FS-complexity). From Figures 5 and 6 one can readily observe that none of the physical properties studied in this work show a uniform enantiomeric behavior, i.e., it is not possible to distinguish the L-aminoacids from the D-ones by using an specific physical property. In contrast, the L-aminoacids can be uniquely characterize d from the D-ones when informatic-theoretical measures are employed (see Figures 7 and 8) and this is perhaps the most interesting result obtained from our work. To the best of our kowledge no similar observations have been reported elsewhere, showing strong evidence of the utility of Information Theory tools for decoding the essential blocks of life.
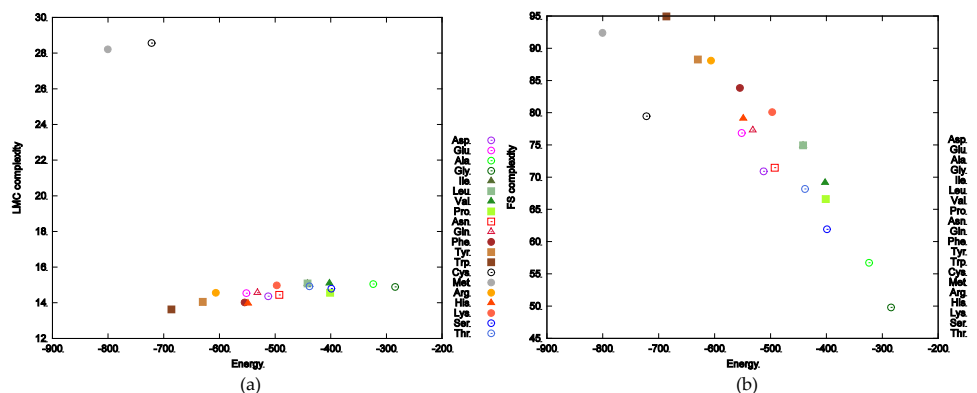
Figure 5. Differences (L-property – D-property) for the energy (in a.u., left) and the ionization potential (in a.u., right) for the set of 20 aminoacids.
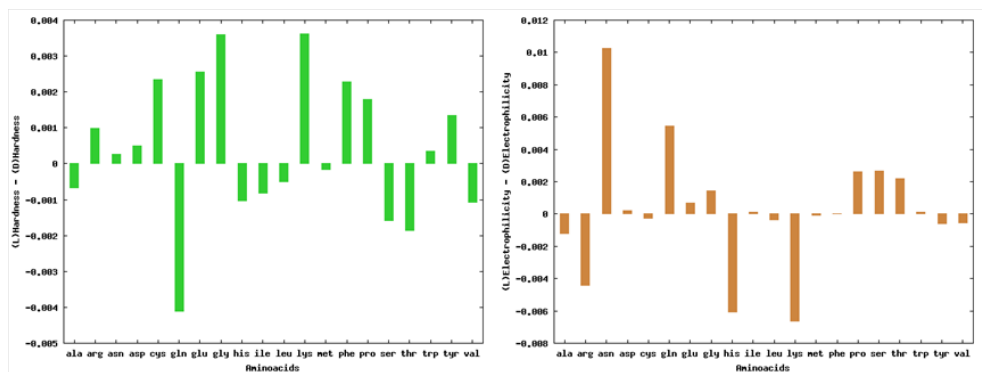
Figure 6. Differences (L-property – D-property) for the hardness (in a.u., left, Eqn. 14) and the electrophilicity (in a.u., right, Eqn. 16) for the set of 20 aminoacids.
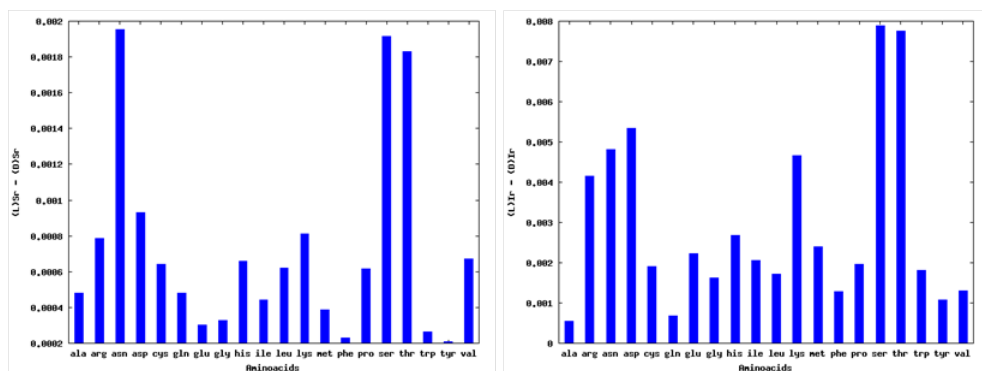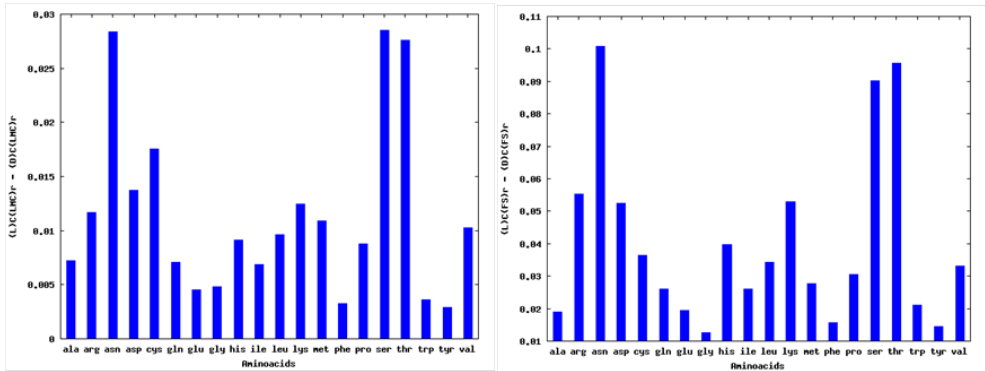

Figure 7. Differences (L-property – D-property) for the Shannon entropy (Eqn. 4, left) and the Fisher information (right, Eqn. 7), both in position space, for the set of 20 aminoacids.

Figure 8. Differences (L-property – D-property) for the LMC (left, Eqn. 2) and the FS-complexities (right, Eqn. 9), both in position space, for the set of 20 aminoacids.

4. Genetic code

4.1. Codons

The genetic code refers to a nearly universal assignment of codons of nucleotides to amino acids. The codon to amino acid assignment is realized through: (i) the code adaptor molecules of transfer RNAs (tRNAs) with a codon's complementary replica (anticodon) and the corresponding amino acid attached to the 3' end, and (ii) aminoacyl tRNA synthetases (aaRSs), the enzymes that actually recognize and connect proper amino acid and tRNAs. The origin of the genetic code is an inherently difficult problem (Crick, 1976). Taking into a count that the events determining the genetic code took place long time ago, and due to the relative compactness of the present genetic code. The degeneracy of the genetic code implies that one or more similar tRNA can recognize the same codon on a messenger mRNA. The number of amino acids and codons is fixed to 20 amino acids and 64 codons (4 nucleotides, A.C.U.G per three of each codon) but the number of tRNA genes varies widely 29 to 126 even between closely related organisms. The frequency of synonymous codon use differs between organisms, within genomes, and along genes, a phenomenon known as CUB (codon usage bias) (Thiele et al., 2011).

Sequences of bases in the coding strand of DNA or in messenger RNA possess coded instructions for building protein chains out of amino acids. There are 20 amino acids used in making proteins, but only four different bases to be used to code for them. Obviously one base can't code for one amino acid. That would leave 16 amino acids with no codes. By taking two bases to code for each amino acid, that would still only give you 16 possible codes (TT, TC, TA, TG, CT, CC, CA and so on) – that is, still not enough. However, by taking three bases per amino acid, that gives you 64 codes (TTT, TTC, TTA, TTG, TCT, TCC and so on). That's enough to

code for everything with lots to spare. You will find a full table of these below. A three base sequence in DNA or RNA is known as a codon.

second base in codon

	U	C	A	G	
U	UUU Phe UUC Phe UUA Leu UUG Leu	UCU Ser UCC Ser UCA Ser UCG Ser	UAU Tyr UAC Tyr UAA stop UAG stop	UGU Cys UGC Cys UGA stop UGG Trp	U C A G
C	CUU Leu CUC Leu CUA Leu CUG Leu	CCU Pro CCC Pro CCA Pro CCG Pro	CAU His CAC His CAA Gln CAG Gln	CGU Arg CGC Arg CGA Arg CGG Arg	U C A G
A	AUU Ile AUC Ile AUA Ile AUG Met	ACU Thr ACC Thr ACA Thr ACG Thr	AAU Asn AAC Asn AAA Lys AAG Lys	AGU Ser AGC Ser AGA Arg AGG Arg	U C A G
G	GUU Val GUC Val GUA Val GUG Val	GCU Ala GCC Ala GCA Ala GCG Ala	GAU Asp GAC Asp GAA Glu GAG Glu	GGU Gly GGC Gly GGA Gly GGG Gly	U C A G

first base in codon (left axis) — *third base in codon* (right axis)

Table 1. Various combinations of three bases in the coding strand of DNA are used to code for individual amino acids - shown by their three letter abbreviation

The codes in the coding strand of DNA and in messenger RNA aren't, of course, identical, because in RNA the base uracil (U) is used instead of thymine (T). Table 1 shows how the various combinations of three bases in the coding strand of DNA are used to code for individual amino acids - shown by their three letter abbreviation. The table is arranged in such a way that it is easy to find any particular combination you want. It is fairly obvious how it works and, in any case, it doesn't take very long just to scan through the table to find what you want. The colours are to stress the fact that most of the amino acids have more than one code. Look, for example, at leucine in the first column. There are six different codons all of which will eventually produce a leucine (Leu) in the protein chain. There are also six for serine (Ser). In fact there are only two amino acids which have only one sequence of bases to code for them - methionine (Met) and tryptophan (Trp). Note that three codons don't have an amino acid but

"stop" instead. For obvious reasons these are known as stop codons. The stop codons in the RNA table (UAA, UAG and UGA) serve as a signal that the end of the chain has been reached during protein synthesis. The codon that marks the start of a protein chain is AUG, that's the amino acid, methionine (Met). That ought to mean that every protein chain must start with methionine.

4.2. Physical and information-theoretical properties

An important goal of the present study is to characterize the biological units which codify aminoacids by means of information-theoretical properties. To accomplished the latter we have depicted in Figures 9 through 13 the Shannon entropy, Disequilibrium, Fisher and the LMC and FS complexities in position space as the number of electron increases, for the group of the 64 codons. A general observation is that all codons hold similar values for all these properties as judging for the small interval values of each graph. For instance, the Shannon entropy values for the aminoacids (see Figure 2) lie between 4.4 to 5.6, whereas the corresponding values for the codons (see Figure 9) lie between 6.66 to 6.82, therefore this information measure serves to characterize all these bilogical molecules, providing in this way the first benchmark informational results for the building blocks of life. Further, it is interesting to note from Figures 9 and 10 that entropy increases with the number of electrons (Fig. 9) whereas the opposite behavior is observed for the Disequilibrium measure. Besides, we may note from these Figures an interesting codification pattern within each isolelectronic group of codons where one may note that an exchange of one nucleotide seems to occur, e.g., as the entropy increases in the 440 electron group the following sequence is found: UUU to (UUC, UCU, CUU) to (UCC, CUC, CCU) to CCC. Similar observations can be obtained from Figures 10 and 11 for D and I, respectively. In particular, Fisher information deserves special analysis, see Figure 11, from which one may observe a more intricated behavior in which all codons seem to be linked across the plot, i.e., note that for each isoelectronic group codonds exchange only one nucleotide, e.g., in the 440 group codons change from UUU to (UUC, UCU, CUU) to (UCC, CUC, CCU) to CCC as the Fisher measure decreaes. Besides, as the Fisher measure and the number of electrons increase linearly a similar exchange is observed, eg., from AAA to (AAG, AGA, GAA) to (AGG, GAG, GGA) to GGG. We believe that the above observations deserve further studies since a codification pattern seems to be apparent.

In Figures 12 and 13 we have depicted the LMC and FS complexities, respectively, where we can note that as the number of electron increases the LMC complexity decreases and the opposite is observed for the FS complexity. It is worth mentioning that similar codification patternsm, as the ones above discussed, are observed for both complexities. Furthermore, we have found interesting to show similar plots in Figures 14 and 15 where the behavior of both complexities is shown with respect to the total energy. It is observed that as the energy increases (negatively) the LMC complexity decreases whereas the FS complexity increases. Note that similar codification patterns are observed in Figure 15 for the FS complexity.
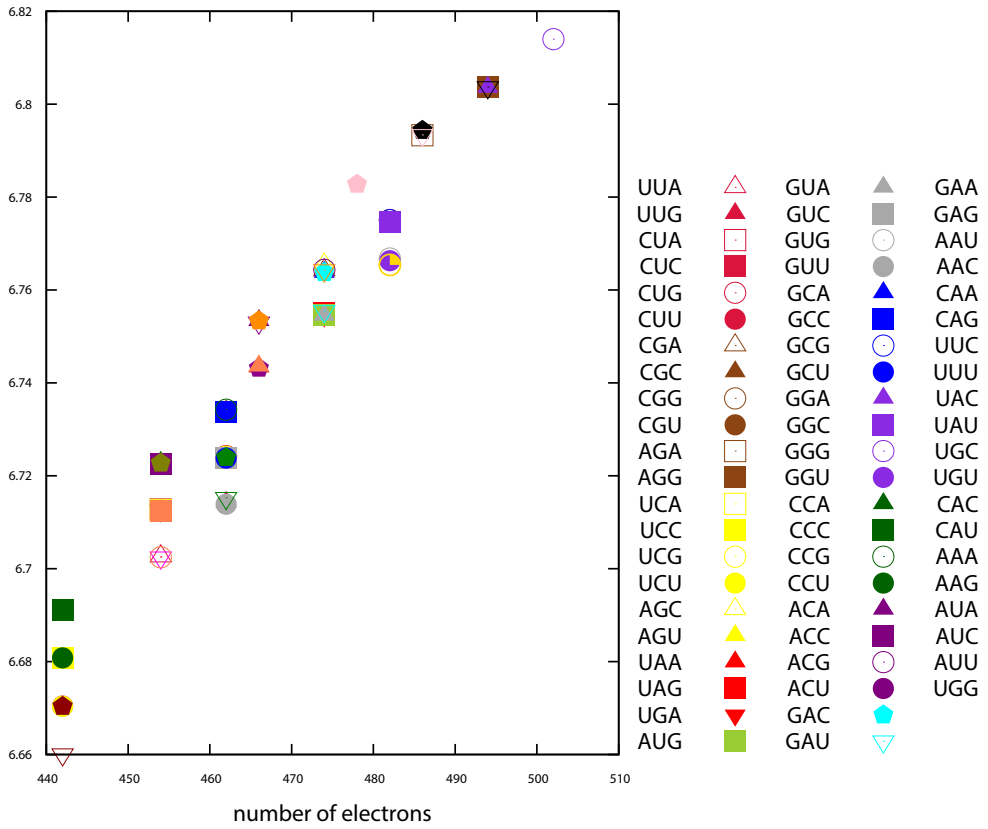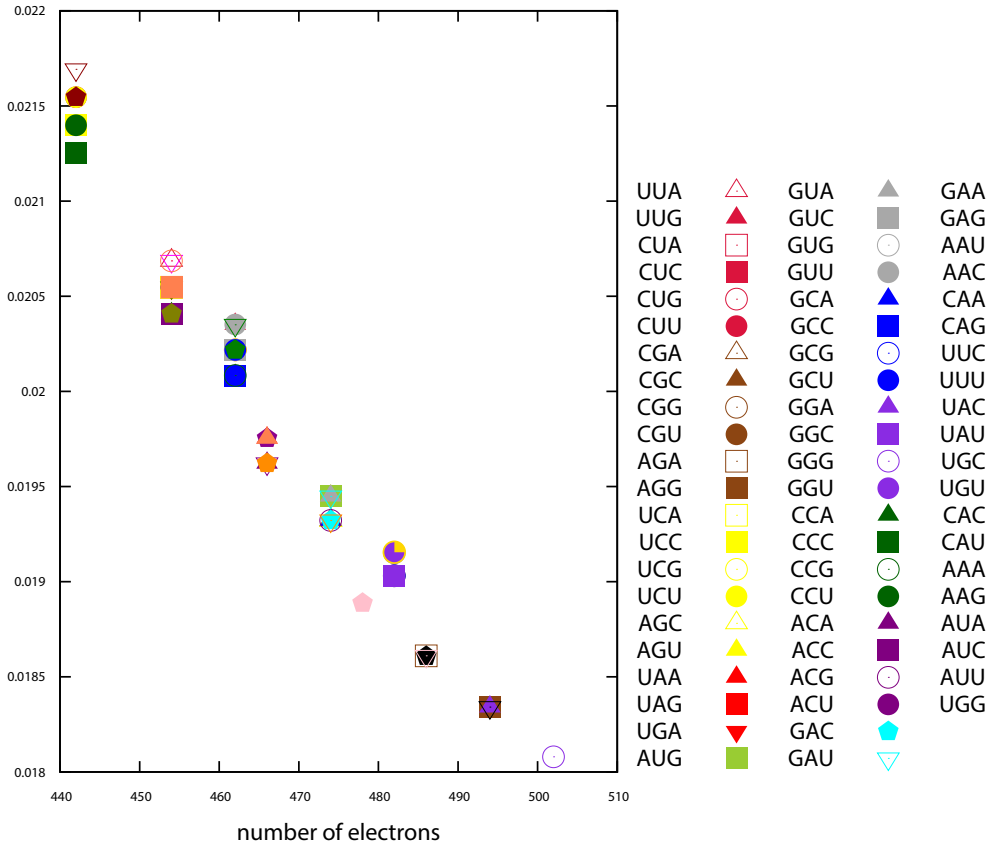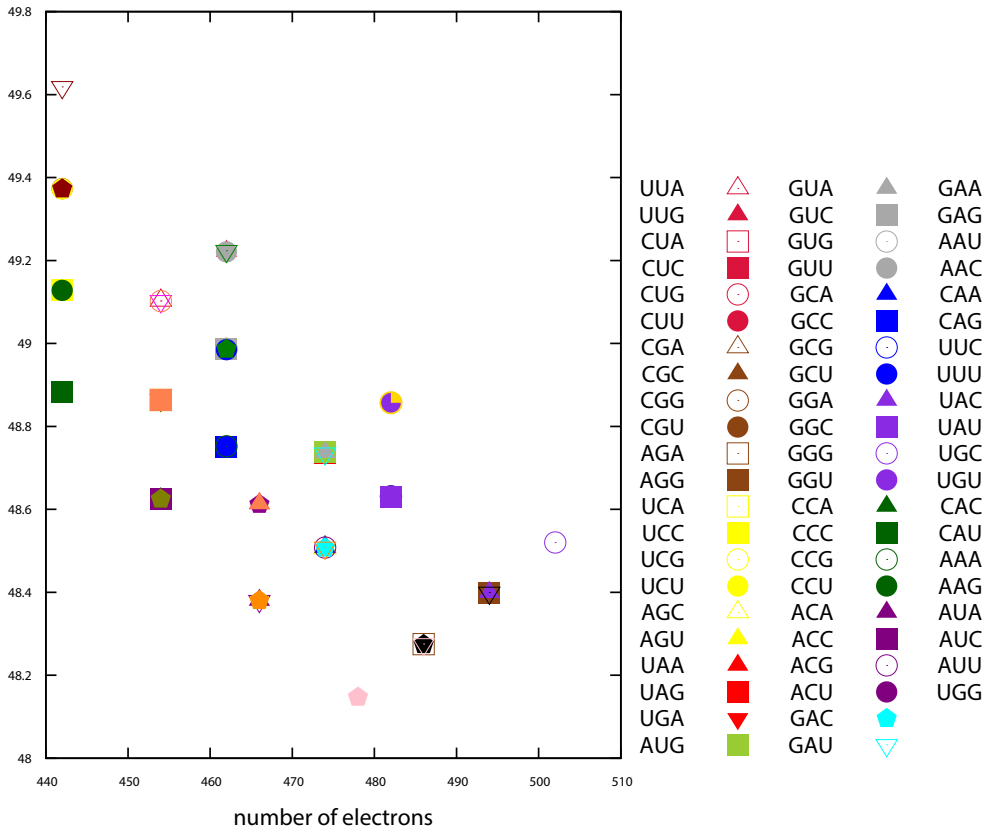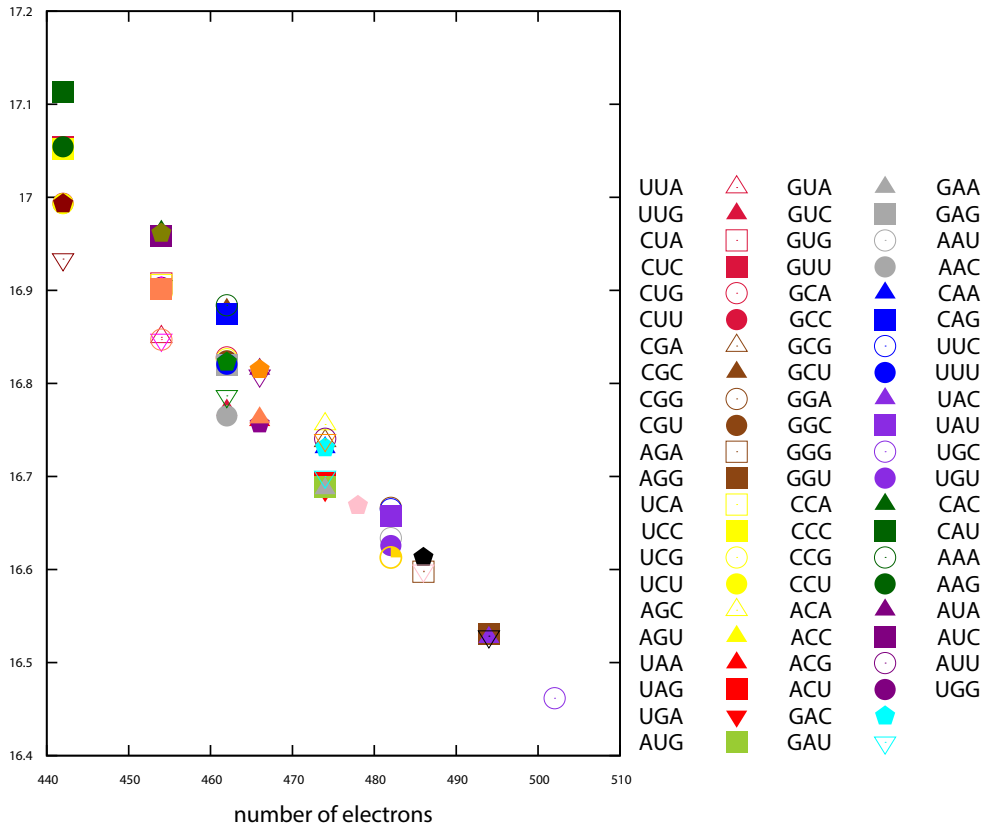
Figure 9. Shannon entropy values in position spaces as a function of the number of electrons for the set of 64 codons.

Figure 10. Disequilibrium in position spaces (Eqn. 3)as a function of the number of electrons for the set of 64 codons
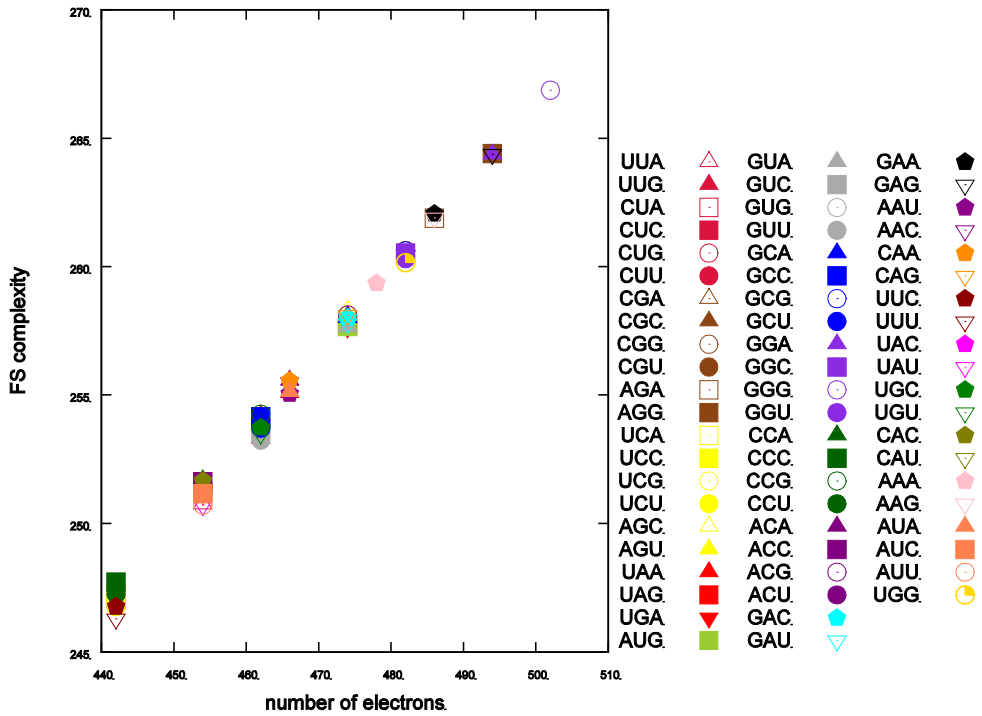
Figure 11. Fisher information in position spaces (Eqn. 7) as a function of the number of electrons for the set of 64 codons

Figure 12. LMC-complexity in position spaces (Eqn.12) as a function of the number of electrons for the set of 64 codons
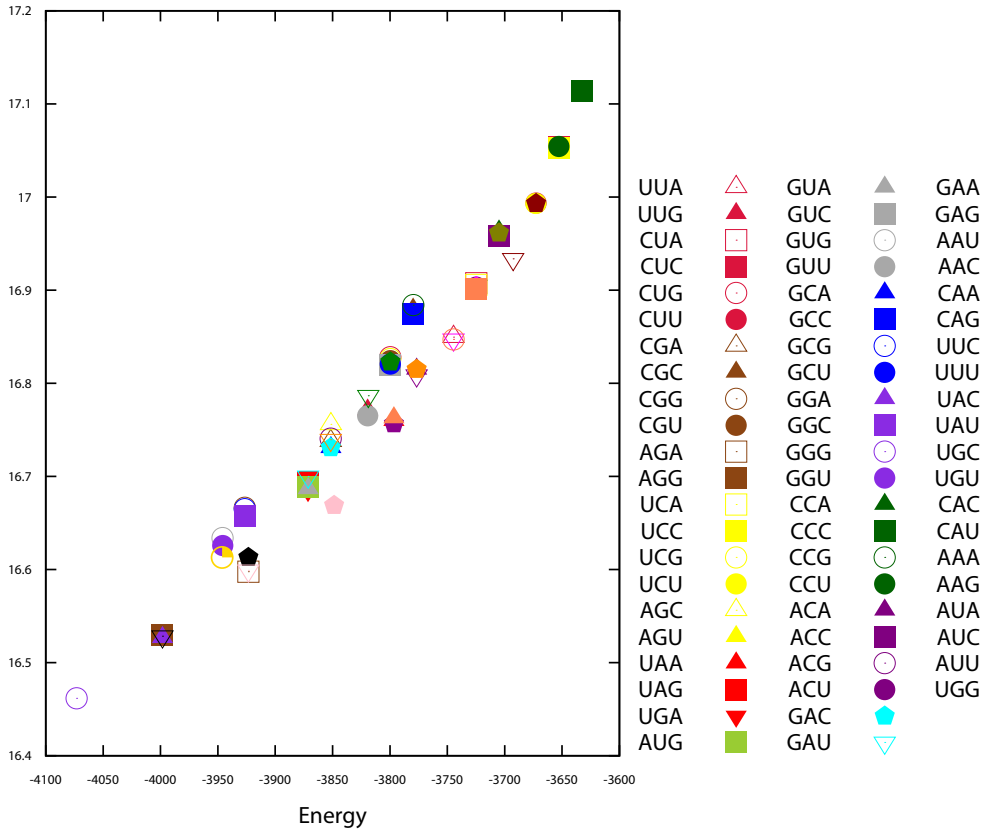
Figure 13. FS-complexity in position spaces (Eqn. 7) as a function of the number of electrons for the set of 64 codons

Figure 14. LMC-complexity in position spaces as a function of the total energy for the set of 64 codons.
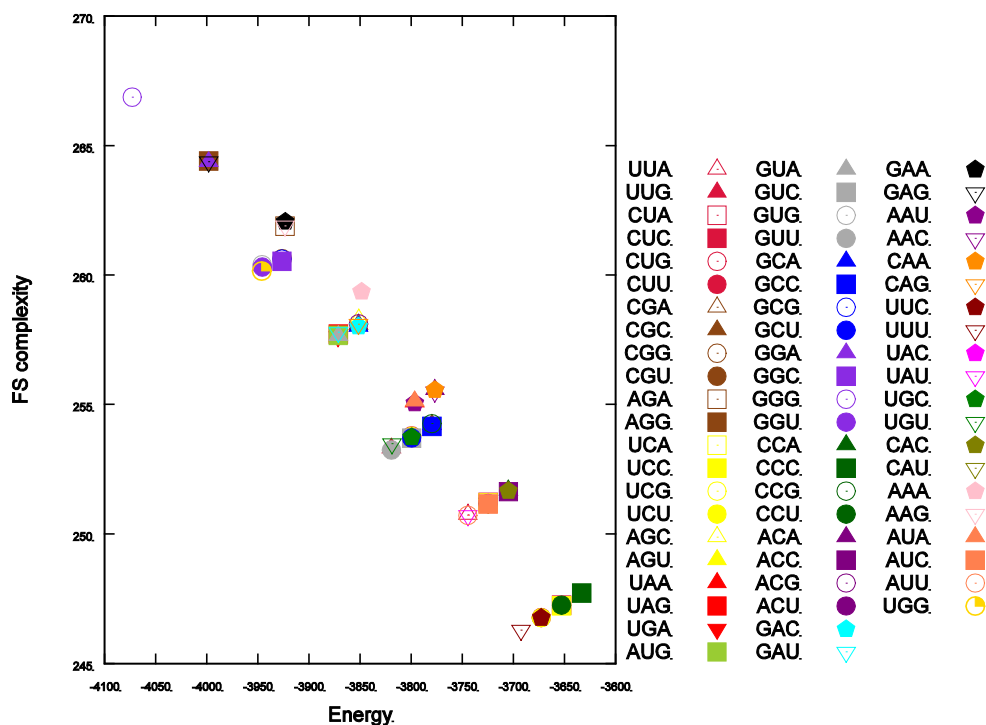
Figure 15. FS-complexity in position space as a function of the total energy for the set of 64 codons.

5. Concluding remarks

We have shown throughout this Chapter that information-theoretical description of the fundamental biological pieces of the genetic code: aminoacids and codons, can be analysed in a simple fashion by employing Information Theory concepts such as local and global information measures and statistical complexity concepts. In particular, we have provided for the first time in the literature with benchmark information-theoretical values for the 20 essential aminoacids and the 64 codons for the nucleotide triplets. Throughout these studies, we believe that information science may conform a new scientific language to explain essential aspects of biological phenomena. These new aspects are not accessible through any other standard methodology in quantum chemistry, allowing to reveal intrincated mechanisms in which chemical phenomena occur. This envisions a new area of research that looks very promising as a standalone and robust science. The purpose of this research is to provide fertile soil to build this nascent scientific area of chemical and biological inquiry through information-theoretical concepts towards the science of the so called Quantum Information Biology.

Acknowledgements

Author details

Rodolfo O. Esquivel[1,6], Moyocoyani Molina-Espíritu[1], Frank Salas[1], Catalina Soriano[2], Carolina Barrientos[3], Jesús S. Dehesa[4,6] and José A. Dobado[5]

1 Departamento de Química, Universidad Autónoma Metropolitana-Iztapalapa, México D.F., México

2 Laboratorio de Química Computacional, FES-Zaragoza, Universidad Nacional Autónoma de México, C.P. 09230 Iztapalapa, México, D.F., México

3 Facultad de Bioanálisis-Veracruz, Universidad Veracruzana, Lab. de Química y Biología Experimental, Veracruz, México

4 Departamento de Física Atómica, Molecular y Nuclear, Universidad de Granada, Granada, Spain

5 Grupo de Modelización y Diseño Molecular, Departamento de Química Orgánica, Universidad de Granada, Granada, Spain

6 Instituto "Carlos I" de Física Teórica y Computacional, Universidad de Granada, Granada, Spain

References

[1]   Angulo, J. C, Antolin, J, Sen, K. D, 2008a, Phys, , & Lett, A.

[2]   Angulo, J. C. ; J, Antolín, , & Chem, J. Phys., 128, 164109 2008

[3] Angulo, J. C, Antolin, J, & Esquivel, R. O. (2010d). Chapter in the monograph'Statisti-
cal Complexity: Applications in Electronic Structure', ed. K.D. Sen (Springer, United
Kingdom, 2010) 978-9-04813-889-0

[4] Anteonodo, C, & Plastino, A. (1996). Phys. Lett. A, 223, 348 1996.

[5] Antolín, J, & Angulo, J. C. Quantum Chem., 109, 586 2009.

[6] Arnett, E. M, & Thompson, O. J. Am. Chem. Soc. , 1981(103), 968-970.

[7] Avalos, M, Babiano, R, Cintas, P, Jiménez, J. L, & Palacios, J. C. (2000). Tetrahedron:
Asymmetry , 2000(11), 2845-874.

[8] Ayers, P. W. ; R. G, Parr, R. G, & Pearson, J. Chem. Phys. 2006, 124, 194107

[9] Bachmann, P. A, Luisi, P. L, & Lang, J. Nature , 1992(357), 57-59.

[10] Bennet, C. H. (1988). The Universal Turing Machine a Half Century, edited by R.
Herhen, Oxford University Press, Oxford, , 227-257.

[11] Betts, M. J, & Russell, R. B. (2003). In Bioinformatics for Geneticists, M.R. Barnes, I.C.
Gray eds, Wiley, 2003

[12] Bombelli, C, & Borocci, S. Lupi, Federica.; Mancini, G.; Mannina, L.; Segre, A. L.; Viel,
Stéphane ((2004). J. Am. Chem. Soc. , 2004(126), 13354-13362.

[13] Borgoo, A, De Proft, F, Geerlings, P, & Sen, K. D. (2007). Chem. Phys. Lett. 444, 186
2007.

[14] Breslow, R. Thetrahedron Letters. (2011). , 2011(52), 4228-4232.

[15] Carbó-dorca, R. ; J, & Arnau, L. Leyda ((1980). Int. J. Quantum Chem., 17, 1185 1980.

[16] Carrera, E. M, Flores-gallegos, N, & Esquivel, R. O. Appl. Math. 233, 1483-1490
(2010).

[17] Catalán, R. G. ; J, & Garay, R. López-Ruiz ((2002). Phys. Rev. E, 66, 011102 2002.

[18] Chaitin, G, 1966, & Acm, J.

[19] Chattaraj, P. K ; U, & Sarkar, D. R. Roy, Chem. Rev., (2006).

[20] Chatzisavvas, K. Ch.; Ch.C. Moustakidis and C.P. Panos ((2005). J. Chem. Phys., 123,
174111 2005

[21] Cleaves, H. J. Biol. , 2010(263), 490-498.

[22] Crick, F, Brenner, S, Klug, A, & Pieczenik, G. (1976). Orig Life 1976, 7:389-397.

[23] Dembo, A, Cover, T. M, & Thomas, J. A. (1991). Information theoretic inequalities",
IEEE Trans. Inform. Theory, , 37, 1501-1518.

[24] Dehesa, J. S. ; R, González-férez, P, Sánchez-moreno, J, & Phys, A.

[25] Esquivel, R O, Flores-gallegos, N, Iuga, C, Carrera, E. M, Angulo, J. C, & Antolín, J. (2009a). Theoretical Chemistry Accounts 124, 445-460 (2009)

[26] Esquivel, R O, & Flores-gallegos, N. Iuga, C; Carrera, E.M. ; Angulo, J.C. ; Antolín J. ((2010a). Physics Letters A , 374(2010), 948-951.

[27] Esquivel, R O, Flores-gallegos, N, Sánchez-dehesa, J. S, Angulo, J. C, Antolín, J, Sen, K, 2010b, Phys,). J, & Chem, A. 2010(114), 1906-1916.

[28] Esquivel, R. O, Liu, S, Angulo, J. C, Dehesa, J. S, Antolin, J, Molina-espíritu, M, 2011a, Phys,). J, Chem, A, Phys, J, & Chem, A.

[29] Esquivel, R. O, Angulo, J. C, Antolin, J, Sánchez-dehesa, J, Flores-gallegos, N, & Lo-pez-rosa, S. Chem. Chem. Phys., 12, 7108-7116 (2010)

[30] Esquivel, R. O, Molina-espíritu, M, Angulo, J. C, Antolín, J, Flores-gallegos, N, Dehe-sa, J. S, & Mol, b. Phys., (2011). , 109

[31] Esquivel, R. O, Flores-gallegos, N, Carrera, E. M, Dehesa, J. S, Angulo, J. C, Antolín, J, & Soriano-correa, C. Sim., 35, (2009), 498-511;

[32] Esquivel, R. O, Flores-gallegos, N, Carrera, E. M, & Soriano-correa, C. (2010g). Jour-nal of Nano Research , 9, 1-15.

[33] Esquivel, R. O, Flores-gallegos, N, Molina-espíritu, M, Plastino, A. R, Dehesa, J. S, Angulo, J. C, Antolín, J, Phys, c). J, & At, B. Mol. Opt. Phys. 44, 175101 ((2011).

[34] Feldman, D. P. ; J. P. Crutchfield (1998)., Phys. Lett. A, 238, 244 (1998).

[35] Fisher, R. A. (1925). Proc. Cambridge Philos. Soc., 22, 700 1925

[36] Flores-gallegos, N, & Esquivel, R. O. Chem. Soc. (2008), 52(1), 19-30

[37] Frieden, B. R. (2004). Science from Fisher Information (Cambridge University Press, Cambridge, 2004.

[38] Fujii, N, Kaji, Y, Fujii, N, Nakamura, T, & Motoie, R. Mori, Yuhei.; Kinouchi, T ((2010). Chemistry & Biodiversity, , 2010(7), 1389-1397.

[39] Geerlings, P. ; F, & De Proft, W. Langenaeker ((2003). Chem. Rev. 2003, 103, 1793.

[40] Ghanty, T. K, & Ghosh, S. K. Chem., 1993, 97, 4951

[41] Hati, S, Datta, D, & Phys, J. Chem., (1994).

[42] He, C. (2010). Nat Chem Biol 2010; , 6, 863-865.

[43] Hein, J. E, & Blackmond, D. G. (2011). Acc. Chem. Res. 201110.1021/ar200316n

[44] Hicks, J. M. (2002). bk- ch001., 2002-0810.

[45] Hornyak, I, & Nagy, A. (2007). Chem. Phys. Lett., 437, 132 2007.

[46] Janak, J. F. (1978). Phys. Rev. B, 1978, 18, 7165.

[47] Kaijser, P. ; V. H. Smith, Jr. ((1997). Adv. Quant. Chem. 10, 37 1997.

[48] Kohout, M. (2007). program DGRID, 2007, version 4.2.

[49] Koopmans, T. A. (1933). Physica, 1933, 1, 104

[50] Kolmogorov, A. N. Probl. Inf. Transm. ((1965).

[51] Lamberti, P. W. ; M. T, Martin, A, & Plastino, O. A. Rosso ((2004). Physica A, 334, 119 2004.

[52] Livingstone, C D, & Barton, G J. (1993). CABIOS, 9 745 1993.

[53] López-rosa, S, Angulo, J. C, Antolín, J, 2009, & Physica, A.

[54] López-rosa, S, & Esquivel, R. O. Angulo, J.C ; Antolin, J.; Dehesa, J.S. ; Flores-Galle-gos, N. ((2010c). J. Chem. Theory Comput. , 6, 145-154.

[55] López-ruiz, R. ; H. L, & Mancini, X. Calbet ((1995). Phys. Lett. A, 1995, 209, 321.

[56] López-ruiz, R. Chem., 115, 215 2005.

[57] Lloyd, S, & Pagels, H. (1988). Ann. Phys. (N.Y.), 188, 186 1988.

[58] Martin, M. T, Plastino, A, & Rosso, O. A. (2003). Phys. Lett. A, 311, 126 2003.

[59] Menger, F. M, & Angelova, M. I. (1998). Acc. Chem. Res. , 1998(31), 789-797.

[60] Nagy, A. Phys., 2003, 119, 9401.

[61] Nagy, A, & Sen, K. D. (2006). Phys. Lett. A, 360, 291 2006.

[62] Nalewajski, R. (2003). Chem. Phys. Lett., 372, 28 2003.

[63] Nanda, V, & Degrado, W. F. Chem. Soc. , 2004(126), 14459-14467.

[64] Onicescu, O. Sci. Paris A 263, 25 1966.

[65] Parr, R. G, & Pearson, R. G. Chem. Soc., 1983, 105, 7512.

[66] Parr, R. G. ; W. Yang, ((1989). Density-Functional Theory of Atoms and Molecules (Oxford University Press: New York, 1989).

[67] Parr, R. G. ; L. V, Szentpály, S, & Liu, J. Am. Chem. Soc., 1999, 121, 1922.

[68] Pearson, R. G. Chem. Soc., 1963, 85, 3533

[69] Pearson, R. G. (1973). Hard and Soft Acids and Bases (Dowen, Hutchinson and Ross: Stroudsberg, 1973)

[70] Pearson, R. G. (1995). Inorg. Chim. Acta 1995, 240, 93.

[71] Pearson, R. G. (1997). Chemical Hardness; (Wiley-VCH; New York, 1997). Ramachan-dran GN, Ramakrishnan C, Sasisekharan J Mol Biol., 7:95-99., 1963

[72] Rawlings, D. C. ; E. R, & Davidson, J. Phys. Chem, 89, 969 1985.

[73] Romera, E. ; J. S, & Dehesa, J. Chem. Phys., 120, 8906 2004.

[74] Romera, E. ; P, Sánchez-moreno, J. S, & Dehesa, J. Math. Phys., 47, 103504 2006

[75] Rosso, O. A, Martin, M. T, Plastino, A, 2003, & Physica, A.

[76] Root-bernstein, R. (2007). BioEssays, , 2007(29), 689-698.

[77] Roy, R. ; A. K, Chandra, S, & Pal, J. Phys. Chem., 1994, 98, 10447

[78] Sen, K. D, Antolín, J, Angulo, J. C, 2007a, Phys, , & Rev, A.

[79] Sen, K. D. ; C. P, & Panos, K. Ch. Chatzisavvas and Ch. Moustakidis ((2007b). Phys. Lett. A, 364, 286 2007.

[80] Shalizi, C. R. ; K. L, & Shalizi, R. Haslinger, ((2004). Phys. Rev. Lett., 93, 118701 2004.

[81] Simon-manso, Y, Fuenteaelba, P, 1998, Phys, J, & Chem, A.

[82] Shannon, C. E, & Weaver, W. (1949). The Mathematical Theory of Communication (University of Illinois Press, Urbana, 1949.

[83] Stryer, L. (1995). Biochemistry, 4th ed., W. H. Freeman and Co.: New York, 1995.

[84] Tamura, K. (2008). BioSystems. , 2008(92), 91-98.

[85] Thiele, I, Fleming, R, Que, R, Bordbar, A, & Palsson, B. (2011). Nature Precedings 2011;

[86] Viedma, C, Ortiz, J. E, Torrea, T, Izumi, T, & Blackmond, D. G. Chem. Soc. , 2008(130), 15274-15275.

[87] Vitanyi, P. M. B, & Li, M. (2000). IEEE Trans. Inf. Theory, 46, 446 2000.

[88] Weber, A. L, & Miller, S. L. Evol. , 1981(17), 273-284.

[89] Werner, F. (2009). Colloids and Surfaces B: Biointerfaces, , 2009(74), 498-503.

[90] Yamano, T, 1995, & Physica, A.

[91] Yamano, T. Phys., 45, 1974 2004.

[92] Zehnacker, A, & Suhm, M. A. (2008). Angew. Chem. Int. Ed. , 2008(47), 6970-6992.