Research Article

Entropy and Multifractality for the Myeloma Multiple TET 2 Gene

Carlo Cattani,[1] Gaetano Pierro,[2] and Giuseppe Altieri[3]

[1] *DipMat, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy*

[2] *System Biology, PhD School, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy*

[3] *SSFO, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy*

Correspondence should be addressed to Carlo Cattani, ccattani@unisa.it

The nucleotide and amino-acid distributions are studied for two variants of mRNA of gene that codes for a protein which is involved in multiple myeloid. Some patches and symmetries are singled out, thus, showing some distinctions between the two variants. Fractal dimensions and entropy are discussed as well.

1. Introduction

In some recent papers, the concepts of fractality [1–19] and entropy [19–21] have been considered as fundamental parameters to investigate the existence of correlations [22–36] and simple rules [37] in the DNA sequences. In particular, it has been observed that the increasing fractal dimension [7–13] can be related to a degeneration in sequences, having as a consequence pathological evolution of related diseases. A fundamental role is played by the concept of information entropy [20, 21] so that a change in the nucleotide distribution in DNA implies a corresponding change in the information content and, as a consequence, a variation in the entropy. Since the cell activity is functionally dependent on the nucleotide distribution our task is to understand better about this distribution and/or about the existence of large scale structure [1–6, 15, 22–37]. So that we could relate the functional activity of cells to some epitomizing patches in the nucleotide distribution. We will propose, in the following, also to take into account the information content in the amino acid distribution. In particular, we will see that the amino-acid distribution shows a higher level structure, and some patchiness which are undetectable in the nucleotide distribution. Our statistical approach is based on the transformation of the symbolic string into a numerical string by the Voss indicator function

[4, 5] which is a discrete binary function. On this function, the indicator matrix is defined and on this matrix the fractal dimension and entropy can be simply computed. We will compare the fractal dimension and complexity of two mRNA variants of TET 2 (ten-eleven translocation 2) gene downloaded from gene bank [38] (similar data are also available from [39–41]), by showing that these parameters can be used to classify the two variants. Multiple myeloma is a pathology which involves plasma cell, but it can move and spread into the whole body. Some aspects are still unclear; however, it is known that this pathology is characterized by the activation of abnormal genes through chromosomal translocations and other genetic anomalies. One of the genes involved in the birth and progression of multiple myeloma is the TET 2. In fact it is present in some myelodysplastic syndromes, and it seems to play a key role when it is subject to mutation. The gene TET 2 is related with myelopoiesis; in fact it encodes protein that we can find significantly expressed in hematopoietic cells and granulocytes.

2. Multiple Myeloma and the Oncogene TET 2

Multiple myeloma (MM) is a blood cancer of the plasma cell. Myeloma originates in a specific type of cell, the plasma cell, but it can move, so that it spreads by the blood to the whole body. Like other cancers, multiple myeloma will develop in steps. Myeloma begins when the normal plasma cell becomes abnormal. The abnormal cell divides, and the new cells divide again and again, thus proliferating the number of abnormal cells. Myeloma cells collect in the bone marrow and in the solid part of the bone. These malignant plasma cells produce a para protein, an inactive antibody known also as M-protein or Bence Jones protein, that attack bone marrow, bones, blood, and kidneys. As a consequence, there happens extensive destruction within the skeleton involving multiple bones, and resulting in widespread bone pain and multiple fractures; for this reason, such a disease is also called multiple myeloma. Some genetic factors are also involved in this pathology. In absence of other symptoms and clinical signs, this condition is more properly called benign monoclonal gammopathy of uncertain significance (MGUS). In fact, the uncertainty about the future progression, it shows that also benign diseases might evolve into MM. It is likely that the evolution of MGUS into MM depends on many mutations of the MGUS clone. Initially, MM has a low progression, but afterward it becomes more aggressive. The signs that characterize onset of multiple myeloma are mostly high concentration of calcium ions with damages in the kidneys, the weakening of the immune system with abnormal production of immune globulin, and some other signs such as an evident osteoporosis. Both MGUS and MM diseases are characterized by the presence of alterations in gene expression [42–46]. The chromosomes that are more involved are 1,11,13,14, respectively. The alteration at chromosome 1 is found in half of cases of MM patients [47–54]. The same aberrations chromosome seem to be evident both in MM and in MGUS, thus supporting the thesis that these two diseases are closely related [53].

Gene TET 2 is located on the chromosome 4 exactly in 4q24. More precisely, the TET2 gene is located from base pair 106,067,942 to base pair 106,200,957 on chromosome 4, position as shown in Figure 1 [38].

The gene TET 2 plays a key role in the conversion of methylcytosine (5mC) to 5-hydroxymethylcytosine (hmC) moreover is related to myelopoiesis. For the hmC many roles were noted like for example (1) remodeling of chromatin structure (2) recruitment of some factors (3) demethylation of cytosine [55, 56]. The gene TET2 encodes a protein that we find significantly expressed in hematopoietic cells and granulocytes. In almost all patients with
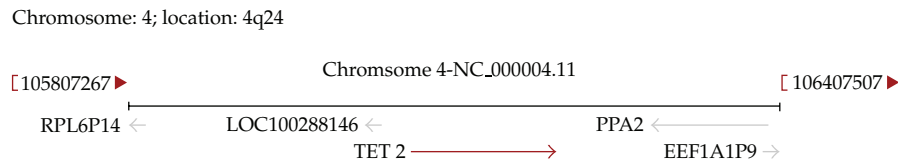
Chromosome: 4; location: 4q24

Figure 1: Location of the TET 2 on chromosome 4.

myelodysplastic syndromes, the protein is decreased in peripheral blood granulocytes. TET 2 gene is usually mutated in myeloproliferative disorders (MPDs). The MPD is part of a larger group of disorders called myeloproliferative neoplasms (MPNs). The mutation of TET 2 characterizes some disorder known as systemic mast cell disease, but TET 2 is above all mutated in myelodisplastic syndromes [57].

We will see that, by using some parameters defined on the indicator function, we can single out some patches which characterize abnormal functional activity [1–3, 35, 58, 59].

3. DNA Representation

The DNA, as well as the mRNA, of each organism of a given species is a sequence of a specific number of base pairs defined on the 4 elements alphabet of nucleotides:

$$A = adenine, \ C = cytosine, \ G = guanine, \ T = thymine. \tag{3.1}$$

Since the base pairs are distributed along a double helix, when straightened, the helix appears as a complementary double-strand system. The two sequences on opposite strands are complementary in the sense that opposite nucleotides must fulfil the ligand rules (A with T and G with C) of base pairs, between purines A and G and pyrimidines T and C. In a DNA sequence, there are some subsequences, which can be roughly subdivided into coding and noncoding regions, having special meaning. In particular, genes (belonging to coding regions) are characteristic sequences of base pairs, and the genes in turn are made by some alternating subsequences of exons and introns (except Procaryotes where the introns are missing). Each exon region is made of triplets of adjacent bases called codons. There are 64 possible codons, inasmuch the number of combination of the 4 nucleotides into 3 length classes. There are only 20 amino acids, therefore, the correspondence codons to amino acids are many to one. The 20 elements alphabet of amino acids is in Table 1. In the following, we will analyze two mRNA sequences: (H1) and (H2), downloaded from the National Center for Biotechnology Information [38], which represent respectively.

(H1) homo sapiens tet oncogene, family member 2 (TET2), transcript variant 1, mRNA, locus NM_001127208 (9796 bp mRNA linear). The accession number is NM_001127208, version NM_001127208.2 GI: 325197189.

(H2) homo sapiens tet oncogene family member 2 (TET2), transcript variant 2, mRNA, locus NM_017628 (9236 bp mRNA linear). The accession number is NM_017628, version NM_017628.4 GI: 325197183.

Some differences between two variants are the following: (H2) is different from (H1) in 5′UTR (untranslate region) and in 3′UTR (untranslate region); furthermore, (H2) variant,

Table 1

	Letter	Name	Codons
1	M	Methionine	ATG
2	E	Glutamic acid	GAA, GAG
3	Q	Glutamine	CAA, CAG
4	D	Aspartic acid	GAT, GAC
5	R	Arginine	CGT, CGC, CGA, CGG, AGA, AGG
6	T	Threonine	ACT, ACC, ACA, ACG
7	N	Asparagines	AAT, AAC
8	H	Histidine	CAT, CAC
9	V	Valine	GTT, GTC, GTA, GTG
10	G	Glycine	GGT, GGC, GGA, GGG
11	L	Leucine	TTA, TTG, CTT, CTC, CTA, CTG
12	S	Serine	TCT, TCC, TCA, TCG, AGT, AGC
13	P	Proline	CCT, CCC, CCA, CCG
14	F	Phenylalanine	TTT, TTC
15	I	Isoleucine	ATT, ATC, ATA
16	C	Cysteine	TGT, TGC
17	A	Alanine	GCT, GCC, GCA, GCG
18	K	Lysine	AAA, AAG
19	Y	Thyroxine	TAT, TAC
20	W	Tryptophan	TGG
		Stop	TAA, TAG, TGA

compared with (H1), is shown to have c-terminal to be distinct and even shorter of (H1), which is also represented by a longer transcript [37].

4. Dot Plot on the Indicator Matrix

In this section, we will define the indicator matrix [4, 5] on which the computation of multifractality and entropy are based.

4.1. Indicator Function for the 4-Symbol Alphabet

Let

$$\aleph_4 = \{A, C, G, T\} \tag{4.1}$$

be the finite set (alphabet) of nucleotides and $x \in \aleph_4$ any member of the 4 symbols alphabet.
A DNA sequence is the finite symbolic sequence

$$\mathfrak{D}(N) = \mathbb{N} \times \aleph_4 \tag{4.2}$$

so that

$$\mathfrak{D}(N) = \{x_h\}_{h=1,\ldots,N}, \quad N < \infty \tag{4.3}$$

being

$$x_h = (h, x) = x(h), \quad (h = 1, 2, \ldots, N; x \in \aleph_4) \tag{4.4}$$

the acid nucleic x at the position h.

Let $\mathfrak{D}_1(N), \mathfrak{D}_2(N)$ be two DNA sequences; the indicator function [4, 5] is the map

$$u : \mathfrak{D}_1(N) \times \mathfrak{D}_2(N) \longrightarrow \{0, 1\} \tag{4.5}$$

such that

$$u_{hk} = u(x_h, x_k) = \begin{cases} 1 & \text{if } x_h = x_k \\ 0 & \text{if } x_h \neq x_k \end{cases} \quad (x_h \in \mathfrak{D}_1(N), x_k \in \mathfrak{D}_2(N)). \tag{4.6}$$

When $\mathfrak{D}_1(N) \equiv \mathfrak{D}_2(N)$ the indicator function, it shows the existence of autocorrelation on the same sequence. According to (4.5), the indicator map of the N-length sequence can be easily represented by the $N \times N$ sparse matrix of binary values $\{0, 1\}$, and this matrix can be visualized by the following (autocorrelation) dot-plot:

$$
\begin{array}{c|ccccccccccc}
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
G & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \ldots \\
C & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & \ldots \\
A & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & \ldots \\
A & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & \ldots \\
T & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \ldots \\
A & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & \ldots \\
C & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & \ldots \\
T & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \ldots \\
G & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \ldots \\
A & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & \ldots \\
\hline
u_{hk} & A & G & T & C & A & T & A & A & C & G & \ldots
\end{array}
\tag{4.7}
$$

4.2. Indicator Function for the 20-Symbols Alphabet of Amino Acids

As a generalization of the 4-symbols alphabet of nucleotides, we can define the 20-symbols alphabet of amino acids as follows:

$$\aleph_{20} = \{M, E, Q, D, R, T, N, H, V, G, L, S, P, F, I, C, A, K, Y, W\}. \tag{4.8}$$

A protein sequence is the finite symbolic sequence

$$\mathfrak{L}(N) = \mathbb{N} \times \aleph_{20} \tag{4.9}$$

so that

$$\mathfrak{L}(N) = \{x_h\}_{h=1,\ldots,N}, \quad N < \infty \tag{4.10}$$

being

$$x_h = (h, x) = x(h), \quad (h = 1, 2, \ldots, N; x \in \aleph_{20}) \tag{4.11}$$

the amino acid x at the position h.

The indicator function [4, 5] can be extended also to protein sequences as the map

$$u : \mathfrak{L}_1(N) \times \mathfrak{L}_2(N) \longrightarrow \{0, 1\} \tag{4.12}$$

such that

$$u_{hk} = u(x_h, x_k) = \begin{cases} 1 & \text{if } x_h = x_k \\ 0 & \text{if } x_h \neq x_k \end{cases} \quad (x_h \in \mathfrak{L}_1(N), x_k \in \mathfrak{L}_2(N)). \tag{4.13}$$

After a transduction of the two DNA sequences (H1) and (H2) into their amino acids components, we can see that the corresponding dot plots can show some (higher-level) structure on the distribution of nucleotides (see **Figure 2**). In particular, (H2) shows a special pattern which is more evident in the amino acids dot plot.

5. Probability Distribution

5.1. Frequency Distribution

The probability distribution of nucleotides can be defined by the frequency

$$p_X(n) = \frac{1}{n} \sum_{i=1}^{n} u_{Xi}, \quad (X \in \aleph_4, \ x_i \in \mathfrak{D}(N); \ 1 \leq n \leq N) \tag{5.1}$$

that the acid nucleic X can be found at the position n. This value can be approximated by the frequency count (on the indicator matrix) of the nucleotide distribution before n. So that, for the transcript variant, we have the probability density distribution of **Figure 3** which, however, tends to assume some different constant values thus showing that nucleotides are heterogeneously distributed.
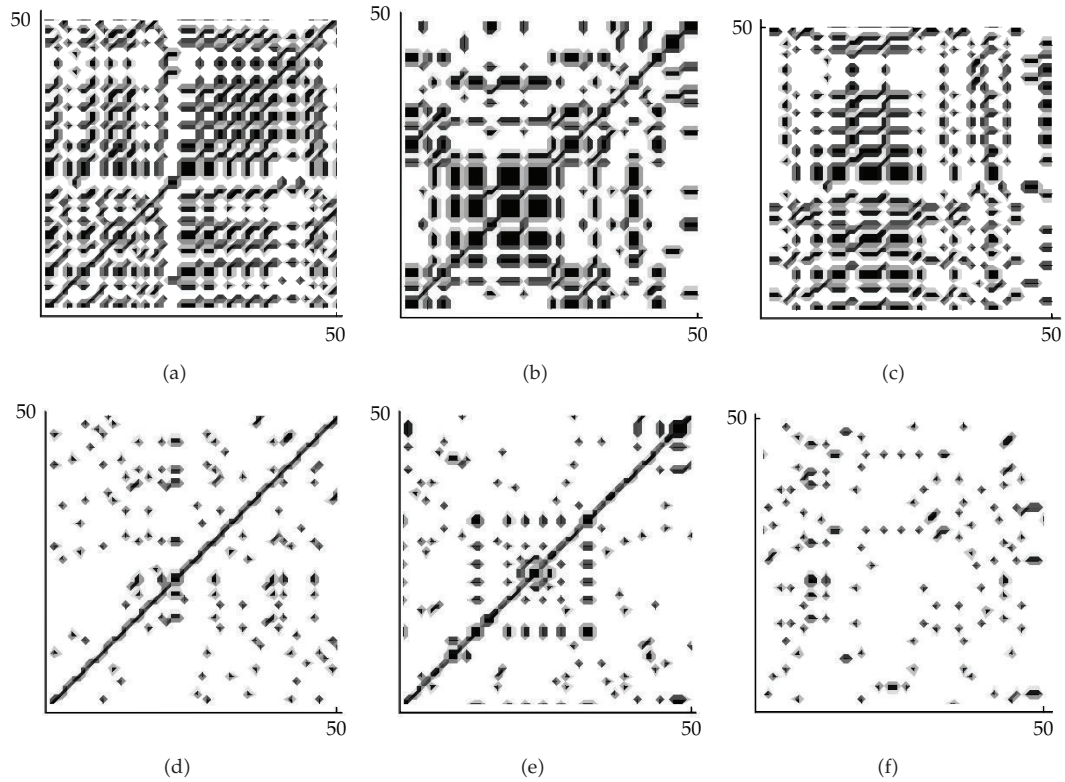
Figure 2: Dot plot for the first 50 nucleotides distribution in the (H1)-(H1), (H2)-(H2) DNA sequences (a,b) and corresponding amino acids (d) and (e). In (c) and (f), the cross correlations (H1)-(H2) are given.

5.2. Distribution of the Essential Amino Acids

Analogously to the nucleotides frequency distribution, we can compute also the amino-acid distribution

$$p_X(n) = \frac{1}{n} \sum_{i=1}^{n} u_{Xi}, \quad (X \in \aleph_{20}, \ x_i \in \mathfrak{L}(N); \ 1 \leq n \leq N) \tag{5.2}$$

that the amino-acid X can be found at the position n.

In particular, we have noticed that even if the nucleotides distribution is nearly the same in both sequences (H1) and (H2), the amino acid shows different distributions for the same amino-acid in each sequence. In other words, the "second"-level distribution seems to be organized according to a different distribution law (see Figures 4 and 5).

6. Fractal Dimension and Entropy

6.1. Fractal Dimension

The frequency distribution implies a corresponding frequency of correlation in the correlation matrix. By using the indicator matrix, it is possible to give a simple formula which enables
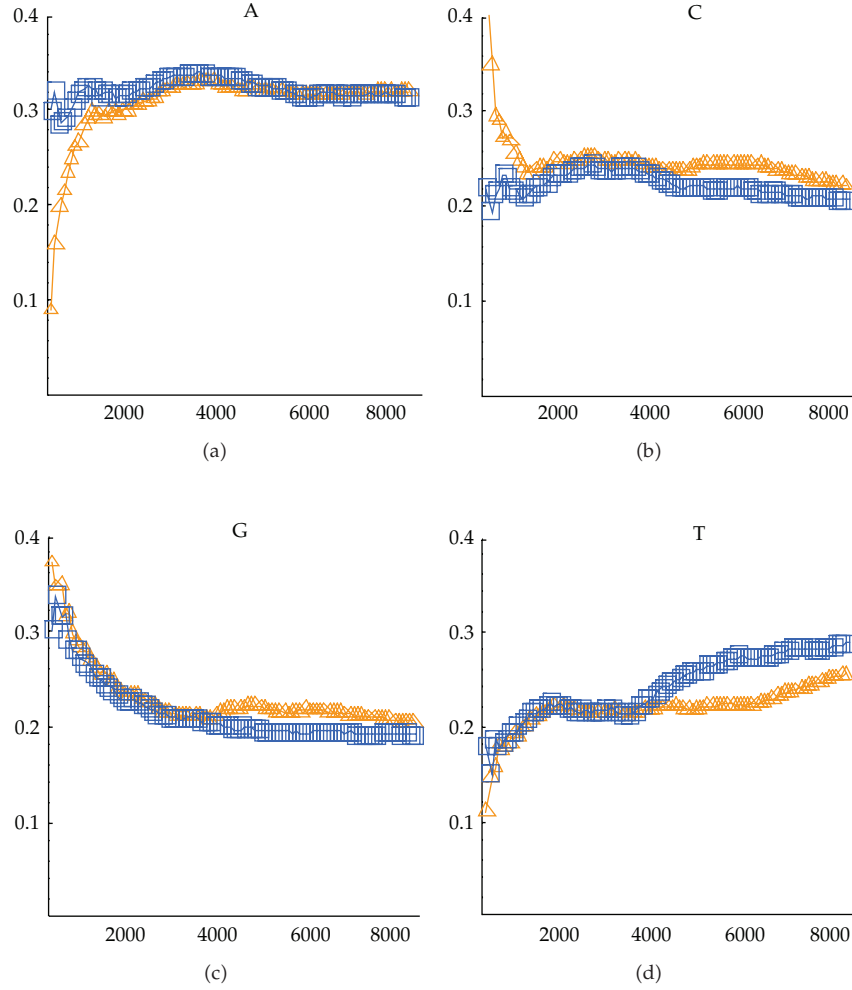
Figure 3: Probability density distribution of nucleotides along TET2 oncogene variants (H1) (brown) and (H2) (blue).

us to estimate the fractal dimension as the average of the number $p(n)$ of 1 in the randomly taken $n \times n$ minors of the $N \times N$ correlation matrix u_{hk}

$$D = \frac{1}{2} \frac{1}{N} \sum_{n=2}^{N} \frac{\log p(n)}{\log n}. \tag{6.1}$$

If we compare the fractal dimensions of the two mRNA sequences (H1) and (H2), we can see (Figure 6) that the fractal dimension of nucleotide distribution tends, for both variants, to the value 1.26.

It is interesting to notice that the corresponding amino acids of the two sequences have (more or less) the same fractal dimension which tends for both (Figure 7) to the value 1.29.
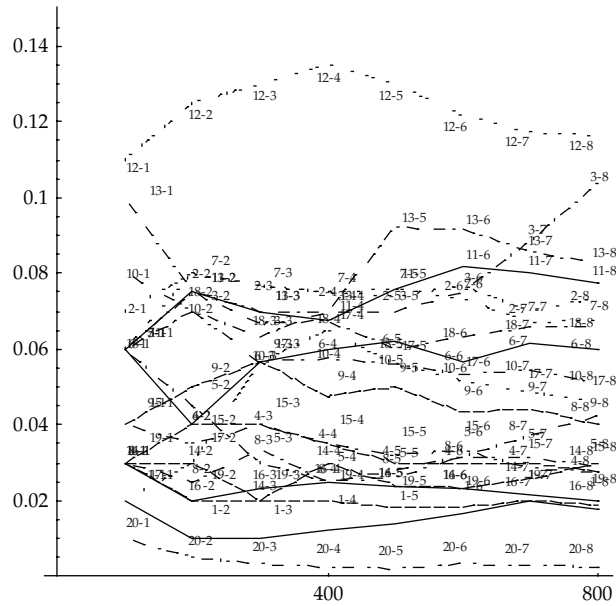
Figure 4: Distribution of amino acids along (H1), being 1 methionine, 2 glutamic acid, 3 glutamine, 4 aspartic acid, 5 arginine, 6 threonine, 7 asparagine, 8 histidine, 9 valine, 10 glycine, 11 leucine, 12 serine, 13 proline, 14 phenylalanine, 15 isoleucine, 16 cysteine, 17 alanine, 18 lysine, 19 thyroxine, and 20 tryptophan.

6.2. Entropy Estimate

As a measure of the information distribution, we consider the normalized Shannon entropy, which is defined, for a distribution over the alphabet $\aleph_\ell$, as

$$
H(n) = -\frac{1}{\log \ell} \sum_{i=1}^{\ell} p_{X_i}(n) \times \begin{cases} \log_2 p_{X_i}(n) & \text{if } p_{X_i}(n) \neq 0, \\ 0 & \text{if } p_{X_i}(n) = 0, \end{cases} \tag{6.2}
$$

where $p_{X_i}(n)$ is given by (5.1) for nucleotides and (5.2) for amino acids.

Since $\sum_{i=1}^{\ell} p_{X_i}(n) = 1$, for all n, the main values of this function are the following.

(1) If $p_{X_i}(n) = 1$, $p_{X_j}(n) = 0$ ($j \neq i$), then $H(n) = 0$. This happens when the information is concentrated in only one symbol.

(2) If $p_{X_i}(n) = p_{X_j}(n) = 1/\ell$, $i \neq j$ then $H(n) = 1$. In this case, the information is equally distributed over all symbols.

(3) Equation $0 \leq H(n) \leq 1$. In general, the information content is distributed over the range $[0, 1]$.

Therefore, the entropy is a positive function ranging in the interval $[0, 1]$, the minimum value is obtained when the distribution is concentrated on a single symbol, while the maximum value is obtained when all symbols are equally distributed.

In particular for higher values of n, according to the frequency definition of probability, the entropy tends to the constant value 1 (see Figures 8 and 9) both for nucleotides and amino
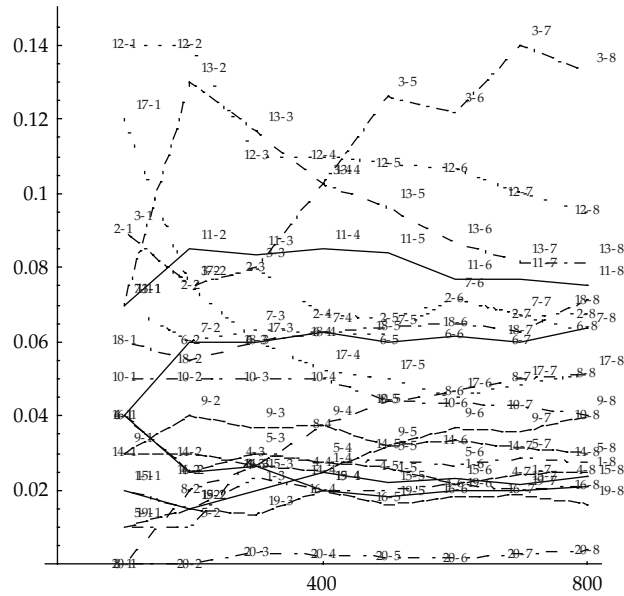
Figure 5: Distribution of amino acids along (H2), being 1 methionine, 2 glutamic acid, 3 glutamine, 4 aspartic acid, 5 arginine, 6 threonine, 7 asparagine, 8 histidine, 9 valine, 10 glycine, 11 leucine, 12 serine, 13 proline, 14 phenylalanine, 15 isoleucine, 16 cysteine, 17 alanine, 18 lysine, 19 thyroxine, and 20 tryptophan.
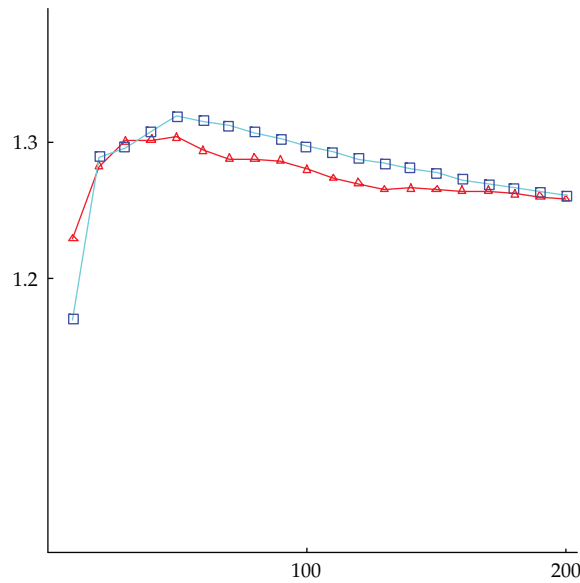

Figure 6: Fractal dimension as function of the length, $i = 10, \ldots, 200$ for (H1) (red) and (H2) (blue).

acids. However, in the first case (Figure 8), the entropy of (H1) is lower than (H2), while on the contrary, for the corresponding amino acids,s the entropy of (H1) is greater than (H2).
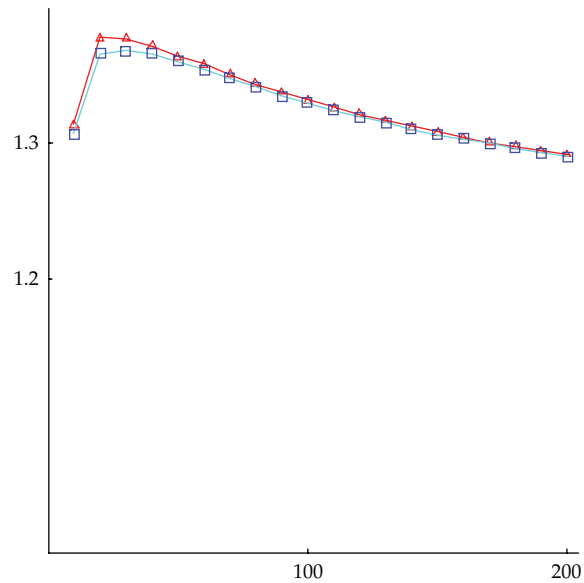
Figure 7: Fractal dimension as function of the length, $i = 10, \ldots, 200$ for the amino acids of (H1) (red) and (H2) (blue).
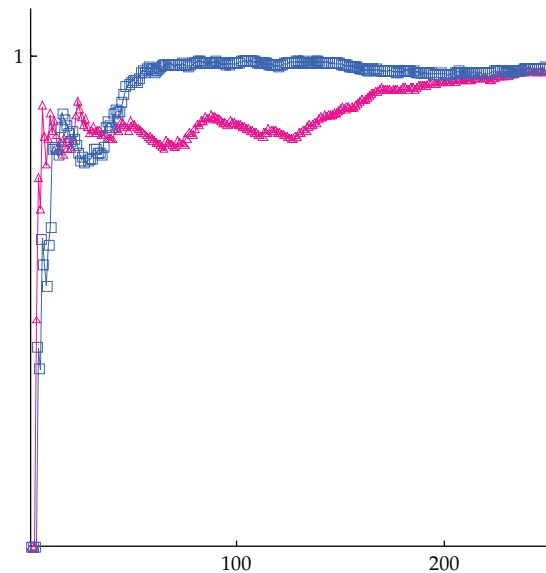

Figure 8: Entropy for the first 300 nucleotides of the sequence (H1) (red) and (H2) (blue).

7. Conclusions

In this paper, two variants of mRNA of isoforms TET2 gene have been analyzed through their nucleotide and amino acids distribution. By using the indicator function (and matrix), the fractal dimension and the entropy have been easily computed. We have noticed that, at
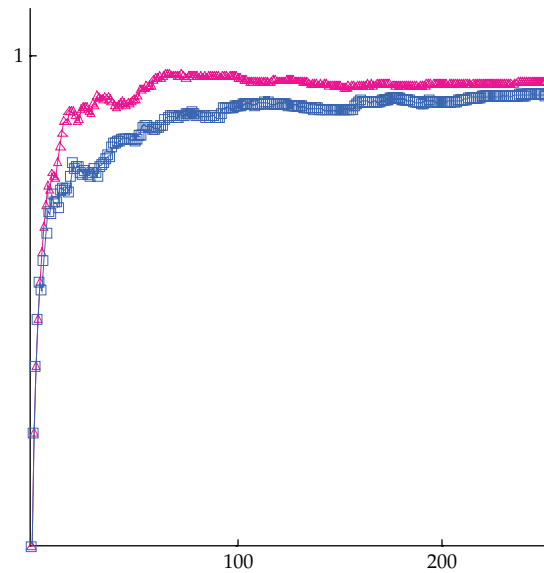
Figure 9: Entropy for the first 300 amino acids of the sequence (H1) (red) and (H2) (blue).

the amino acid level, some patches can be easily singled out. Moreover, the second variant (H2) of TET 2 shows some more randomness than (H1).

References

[1] C. Cattani, "Fractals and hidden symmetries in DNA," *Mathematical Problems in Engineering*, vol. 2010, Article ID 507056, 31 pages, 2010.

[2] C. Cattani, "Wavelet algorithms for DNA analysis," in *Computational Molecular Biology: Techniques, Approaches and Applications*, M. Elloumi and A. Y. Zomaya, Eds., Wiley Series in Bioinformatics, chapter 35, John Wiley & Sons, New York, NY, USA, 2010.

[3] C. Cattani and G. Pierro, "Complexity on acute myeloid leukemia mRNA transcript variant," *Mathematical Problems in Engineering*, vol. 2011, Article ID 379873, 16 pages, 2011.

[4] R. F. Voss, "Evolution of long-range fractal correlations and 1/f noise in DNA base sequences," *Physical Review Letters*, vol. 68, no. 25, pp. 3805–3808, 1992.

[5] R. F. Voss, "Long-Range Fractal Correlations in DNA introns and exons," *Fractals*, vol. 2, pp. 1–6, 1992.

[6] S. V. Buldyrev, A. L. Goldberger, S. Havlin et al., "Long-range correlation properties of coding and noncoding DNA sequences: GenBank analysis," *Physical Review E*, vol. 51, no. 5, pp. 5084–5091, 1995.

[7] K. Metze, "Fractal dimension of chromatin and cancer prognosis," *Epigenomics*, vol. 2, no. 5, pp. 601–604, 2010.

[8] R. L. Adam, R. C. Silva, F. G. Pereira, N. J. Leite, I. Lorand-Metze, and K. Metze, "The fractal dimension of nuclear chromatin as a prognostic factor in acute precursor B lymphoblastic leukemia," *Cellular Oncology*, vol. 28, no. 1-2, pp. 55–59, 2006.

[9] K. Metze, I. Lorand-Metze, N. J. Leite, and R. L. Adam, "Goodness-of-fit of the fractal dimension as a prognostic factor," *Cellular Oncology*, vol. 31, no. 6, pp. 503–504, 2009.

[10] L. Goutzanis, N. Papadogeorgakis, P. M. Pavlopoulos et al., "Nuclear fractal dimension as a prognostic factor in oral squamous cell carcinoma," *Oral Oncology*, vol. 44, no. 4, pp. 345–353, 2008.

[11] A. Mashiah, O. Wolach, J. Sandbank, O. Uziel, P. Raanani, and M. Lahav, "Lymphoma and leukemia cells possess fractal dimensions that correlate with their biological features," *Acta Haematologica*, vol. 119, no. 3, pp. 142–150, 2008.

[12] K. Metze, D. P. Ferro, M. A. Falconi et al., "Fractal characteristics of nuclear chromatin in routinely stained cytology are independent prognostic factors in patients with multiple myeloma," *Virchows Archiv*, vol. 445, supplement 1, pp. 7–21, 2009.

[13] V. Bedin, R. L. Adam, B. C. S. de Sá, G. Landman, and K. Metze, "Fractal dimension of chromatin is an independent prognostic factor for survival in melanoma," *BMC Cancer*, vol. 10, article 260, 2010.

[14] D. V. Lebedev, M. V. Filatov, A. I. Kuklin et al., "Fractal nature of chromatin organization in interphase chicken erythrocyte nuclei: DNA structure exhibits biphasic fractal properties," *FEBS Letters*, vol. 579, no. 6, pp. 1465–1468, 2005.

[15] J. G. McNally and D. Mazza, "Fractal geometry in the nucleus," *The EMBO Journal*, vol. 29, no. 1, pp. 2–3, 2010.

[16] M. Takahashi, "A fractal model of chromosomes and chromosomal DNA replication," *Journal of Theoretical Biology*, vol. 141, no. 1, pp. 117–136, 1989.

[17] A. Delides, I. Panayiotides, A. Alegakis et al., "Fractal dimension as a prognostic factor for laryngeal carcinoma," *Anticancer Research*, vol. 25, no. 3, pp. 2141–2144, 2005.

[18] R. C. Ferreira, P. S. de Matos, R. L. Adam, N. J. Leite, and K. Metze, "Application of the Minkowski-Bouligand fractal dimension for the differential diagnosis of thyroid follicular neoplasias," *Cellular Oncology*, vol. 28, no. 5-6, pp. 331–333, 2006.

[19] J. M. Adams, A. W. Harris, and C. A. Pinkert, "The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice," *Nature*, vol. 318, no. 6046, pp. 533–538, 1985.

[20] L. Pontrjagin and L. Schnirelmann, "Sur une propriété métrique de la dimension," *Annals of Mathematics*, vol. 33, pp. 156–162, 1932.

[21] A. N. Kolmogorov and V. M. Tihomiroff, "ε_{λ}-Entropy and ε-capacity of sets in functional spaces," *Uspekhi Mat. Nauk*, vol. 14, no. 2, pp. 3–86, 1949.

[22] J. Patrick Fitch and B. Sokhansanj, "Genomic engineering: moving beyond DNA sequence to function," *Proceedings of the IEEE*, vol. 88, no. 12, pp. 1949–1971, 2000.

[23] H. Gee, "A journey into the genome: what's there," *Nature*, 2001.

[24] P. D. Cristea, "Large scale features in DNA genomic signals," *Signal Processing*, vol. 83, no. 4, pp. 871–888, 2003.

[25] H. Herzel, E. N. Trifonov, O. Weiss, and I. Große, "Interpreting correlations in biosequences," *Physica A*, vol. 249, no. 1–4, pp. 449–459, 1998.

[26] W. Li, "The study of correlation structures of DNA sequences: a critical review," *Computers and Chemistry*, vol. 21, no. 4, pp. 257–271, 1997.

[27] W. Li and K. Kaneko, "Long-range correlations and partial spectrum in a noncoding DNA sequence," *Europhysics Letters*, vol. 17, pp. 655–660, 1992.

[28] C. K. Peng, S. V. Buldyrev, A. L. Goldberger et al., "Long-range correlations in nucleotide sequences," *Nature*, vol. 356, no. 6365, pp. 168–170, 1992.

[29] C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, "Mosaic organization of DNA nucleotides," *Physical Review E*, vol. 49, no. 2, pp. 1685–1689, 1994.

[30] O. Weiss and H. Herzel, "Correlations in protein sequences and property codes," *Journal of Theoretical Biology*, vol. 190, no. 4, pp. 341–353, 1998.

[31] Z. G. Yu, V. V. Anh, and B. Wang, "Correlation property of length sequences based on global structure of the complete genome," *Physical Review E*, vol. 63, no. 1, Article ID 011903, 2001.

[32] P. P. Vaidyanathan and B. J. Yoon, "The role of signal-processing concepts in genomics and proteomics," *Journal of the Franklin Institute*, vol. 341, no. 1-2, pp. 111–135, 2004.

[33] P. Bernaola-Galván, R. Román-Roldán, and J. L. Oliver, "Compositional segmentation and long-range fractal correlations in DNA sequences," *Physical Review E*, vol. 53, no. 5, pp. 5181–5189, 1996.

[34] W. Li and J. Bentley, "The complexity of DNA: the measure of compositional heterogenity in DNA sequence and measures of complexity," *Complexity*, vol. 3, pp. 33–37, 1997.

[35] S. Karlin and V. Brendel, "Patchiness and correlations in DNA sequences," *Science*, vol. 259, no. 5095, pp. 677–680, 1993.

[36] J. M. Bennett, M. L. Young, J. W. Andersen et al., "Long-term survival in acute myeloid leukemia: the Eastern Cooperative Oncology Group experience," *Cancer*, vol. 80, no. 11, pp. 2205–2209, 1997.

[37] E. Schrödinger, *What is Life? Physical Aspects of Living Cell*, Cambridge University Press, Cambridge, UK, 1948.

[38] National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/genbank/.

[39] Universal Protein Resource, http://www.uniprot.org/help/about.

[40] Gene Location, Weizmann Institute of Science, http://genecards.weizmann.ac.il/geneloc/.

[41] e!Ensemble, Ensembl project, EMBL-EBI & Wellcome Trust Sanger Institute, http://www.ensemble.org.

[42] H. Avet-Loiseau, T. Facon, A. Daviet et al., "14q32 translocations and monosomy 13 observed in monoclonal gammopathy of undetermined significance delineate a multistep process for the oncogenesis of multiple myeloma," *Cancer Research*, vol. 59, no. 18, pp. 4546–4550, 1999.

[43] J. Drach, J. Schuster, H. Nowotny et al., "Multiple myeloma: high incidence of chromosomal aneuploidy as detected by interphase fluorescence in situ hybridization," *Cancer Research*, vol. 55, no. 17, pp. 3854–3859, 1995.

[44] M. Flactif, M. Zandecki, J. L. Laï et al., "Interphase fluorescence in situ hybridization (FISH) as a powerful tool for the detection of aneuploidy in multiple myeloma," *Leukemia*, vol. 9, no. 12, pp. 2109–2114, 1995.

[45] R. Fonseca, G. J. Ahmann, S. M. Jalal et al., "Chromosomal abnormalities in systemic amyloidosis," *British Journal of Haematology*, vol. 103, no. 3, pp. 704–710, 1998.

[46] M. Zandecki, J.-L. Laï, F. Geneviève et al., "Several cytogenetic subclones may be identified within plasma cells from patients with monoclonal gammopathy of undetermined significance, both at diagnosis and during the indolent course of this condition," *Blood*, vol. 90, no. 9, pp. 3682–3690, 1997.

[47] G. W. Dewald, R. A. Kyle, G. A. Hicks, and P. R. Greipp, "The clinical significance of cytogenetic studies in 100 patients with multiple myeloma, plasma cell leukemia, or amyloidosis," *Blood*, vol. 66, no. 2, pp. 380–390, 1985.

[48] B. Barlogie, J. Epstein, P. Selvanayagam, and R. Alexanian, "Plasma cell myeloma—New biological insights and advances in therapy," *Blood*, vol. 73, no. 4, pp. 865–879, 1989.

[49] W. Liang, J. E. Hopper, and J. D. Rowley, "Karyotypic abnormalities and clinical aspects of patients with multiple myeloma and related paraproteinemic disorders," *Cancer*, vol. 44, no. 2, pp. 630–644, 1979.

[50] G. Gahrton, L. Zech, and K. Nillsson, "2 Translocations, t(11;14) and t(1;6), in a patient with plasma cell leukaemia and 2 populations of plasma cells," *Scandinavian Journal of Haematology*, vol. 24, no. 1, pp. 42–46, 1980.

[51] G. J. Morgan, F. E. Davies, and M. Linet, "Myeloma aetiology and epidemiology," *Biomedicine and Pharmacotherapy*, vol. 56, no. 5, pp. 223–234, 2002.

[52] H. Avet-Loiseau, M. Attal, P. Moreau et al., "Genetic abnormalities and survival in multiple myeloma: the experience of the Intergroupe Francophone du Myélome," *Blood*, vol. 109, no. 8, pp. 3489–3495, 2007.

[53] J. R. Sawyer, J. A. Waldron, S. Jagannath, and B. Barlogie, "Cytogenetic findings in 200 patients with multiple myeloma," *Cancer Genetics and Cytogenetics*, vol. 82, no. 1, pp. 41–49, 1995.

[54] O. Landgren, R. A. Kyle, R. M. Pfeiffer et al., "Monoclonal gammopathy of undetermined significance (MGUS) consistently precedes multiple myeloma: a prospective study," *Blood*, vol. 113, no. 22, pp. 5412–5417, 2009.

[55] S. M. C. Langemeijer, R. P. Kuiper, M. Berends et al., "Acquired mutations in TET2 are common in myelodysplastic syndromes," *Nature Genetics*, vol. 41, no. 7, pp. 838–842, 2009.

[56] M. Ko, Y. Huang, A. M. Jankowska et al., "Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2," *Nature*, vol. 468, no. 7325, pp. 839–843, 2010.

[57] R. B. Lorsback, J. Moore, S. Mathew, S. C. Raimondi, S. T. Mukatira, and J. R. Downing, "TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;23) [3]," *Leukemia*, vol. 17, no. 3, pp. 637–641, 2003.

[58] S. Y. Chen and Q. Guan, "Parametric shape representation by a deformable NURBS model for cardiac functional measurements," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, part 1, pp. 480–487, 2011.

[59] S. Chen, J. Zhang, H. Zhang et al., "Myocardial motion analysis for determination of tei-index of human heart," *Sensors*, vol. 10, no. 12, pp. 11428–11439, 2010.