What can we learn with wavelets about DNA sequences?

Alain Arneodo, Yves d'Aubenton-Carafa, Benjamin Audit, Emmanuel Bacry, Jean-François Muzy, Claude Thermes

What can we learn with wavelets about DNA sequences?

A. Arneodo [a,], Y. D'Aubenton-Carafa [b], B. Audit [a], E. Bacry [c], J.F. Muzy [a], C. Thermes [b]

[a] *Centre de Recherche Paul Pascal, Avenue Schweitzer, 33600 Pessac, France*
[b] *Centre de Génétique Moléculaire du CNRS, Laboratoire Associé à l'Université Pierre et Marie Curie, Allée de la Terrasse, 91198 Gif-sur-Yvette, France*
[c] *Centre de Mathématiques Appliquées, Ecole Polytechnique, 91128 Palaiseau, France*

We use the wavelet transform to explore the complexity of DNA sequences. Long-range correlations are clearly identified and shown to be related to the sequence GC content. The significance of this observation to gene evolution is discussed.

Keywords: Wavelet transform; Fractal scaling; Multifractal formalism; Long-range correlations; Fractional Brownian motions; DNA sequences; Isochores

---

1. Introduction

The possible relevance of scale invariance and fractal concepts to the structural complexity of genomic sequences is the subject of considerable increasing interest [1]. During the past few years, there has been intense discussion about the existence, the nature and the origin of long-range correlations in DNA sequences. Different techniques including mutual information functions [2], autocorrelation functions [3,4], power spectra [5,6], "DNA walk" representation [1,7], Zipf analysis [8] were used for statistical analysis of DNA sequences. But despite the effort spent, there is still some continuing debate on rather struggling questions. In that respect, it is of fundamental importance to corroborate the fact that the reported long-range correlations are not just an artefact of the compositional heterogeneity of the genome organization [3,4,9–12]. Furthermore, it is still an open question whether the long-range correlation properties are different for protein-coding (exonic) and noncoding (intronic, intergenetic) sequences [1–8,13]. One of the main obstacles to long-range correlation analysis is the mosaic structure of

DNA sequences which are well known to be formed of "patches" ("strand bias") of different underlying compositions [14–16]. These patches appear as trends in the DNA walk landscapes and are likely to introduce some breaking of scale invariance [9–12]. Most of the technique used so far for characterizing the presence of long-range correlations are not well adapted to study patchy sequences. In a preliminary work [17,18], we have emphasized the wavelet transform (WT) as a very powerful technique for fractal analysis of DNA sequences. By considering analyzing wavelets that make the WT microscope blind to low-frequency trends, one can reveal and quantify the scaling properties of DNA walks. Here we report on recent results obtained by applying the so-called wavelet transform modulus maxima (WTMM) method [19,20] to various genomic sequences mainly selected in the human genome.

2. A wavelet-based multifractal formalism

In order to characterize the singular behavior of a distribution f at a point x_o, one can use the Hölder (or roughness) exponent $h(x_o)$ which basically corresponds to the highest exponent h which satisfies: $|f(x) - P(x)| \sim |x - x_o|^h$ $(x \to x_o)$, where the polynomial $P(x)$ is of order smaller than h. The so-defined exponent $h(x_o)$ "quantifies" the singularity strengh of f at x_o: the higher $h(x_o)$, the less singular f around x_o. The singularity spectrum $D(h)$ is then defined as the Hausdorff dimension of the set of points x where the Hölder exponent of f is h. It thus gives a statistical characterization of the different singular behavior involved in f. In order to estimate locally the Hölder exponent, one needs a tool which must be blind to local smooth behavior, i.e., to the polynomial term $P(x)$. The WT [19,20] is perfectly adapted to such requirements. It is a space-scale analysis which consists in expanding distributions in terms of wavelets which are constructed from a single function, the analyzing wavelet ψ, by means of dilatations and translations. The WT is defined as

$$T_\psi[f](x_o, a) = \frac{1}{a} \int_{-\infty}^{+\infty} f(x) \psi\left(\frac{x - x_o}{a}\right) dx \,, \tag{1}$$

where x_o is the space parameter and a (> 0) the scale parameter. By choosing ψ so that its first N moments are zero, one can easily prove [19,20] that provided $N > h(x_o)$, $h(x_o)$ can be obtained locally from the power-law behavior of the WT, $T_\psi(x_o, a) \sim a^{h(x_o)}$, in the limit $a \to 0^+$ (we discard here the possible existence of oscillating singularities). In this work, we will mainly use the derivatives of the Gaussian function as analyzing wavelets: $\psi^{(N)} = (d^N/dx^N)(e^{-x^2/2})$.

A natural way of performing a multifractal analysis of fractal functions consists in generalizing classical box-counting techniques by using wavelets instead of boxes. The WTMM method [19,20] amounts to investigate the scaling behavior of some partition functions defined in terms of wavelet coefficients:

$$Z(q, a) = \sum_{x_i \in \mathscr{S}(a)} |T_\psi(x_i, a)|^q \sim a^{\tau(q)} \,, \tag{2}$$

where $q \in \mathbb{R}$. The sum is taken over the WT skeleton $\mathscr{S}(a)$ defined, at each scale a, by the set of all the points x_i that correspond to local maxima of $|T_\psi(x,a)|$ considered as a function of x. The main result of the WTMM method is that the $D(h)$ singularity spectrum can be determined from the Legendre transform of the scaling exponent $\tau(q)$: $D(h) = \min_q(qh - \tau(q))$. Homogeneous fractal functions that involve singularities of unique Hölder exponent $h(x) = h$, are characterized by a linear $\tau(q)$ spectrum ($h = \partial\tau/\partial q$). On the contrary, a nonlinear $\tau(q)$ curve is the signature of nonhomogeneous functions that display multifractal properties (i.e., $h(x)$ is a fluctuating quantity that depends upon x). For its ability to resolve multifractal scaling via the estimate of the entire $\tau(q)$ spectrum, the WTMM method [17,18] is a definite step beyond the technique used so far in the literature [1–7] which were restricted to the estimate of the second-order exponent $\tau(2)$ only. The reliability of the WTMM method has been tested on various mathematical examples including fractional Brownian motions (fBm's). The fBm's $B_H(x)$ are Gaussian stochastic processes of zero mean with stationary increments. They are indexed by a parameter $H(0 < H < 1)$ that accounts for the presence ($H \neq \frac{1}{2}$) or the absence ($H = \frac{1}{2}$) of correlations between increments. The fBm's are statistically homogeneous fractals characterized by a single Hölder exponent $h = H$ and thus by a linear $\tau(q)$ spectrum: $\tau(q) = qH - 1$. The WTMM method has already been successfully applied to numerical and experimental data from various domains such as fully developed turbulence and fractal growth phenomena [19,20].

3. Wavelet analysis of DNA sequences

3.1. How to make the WT microscope blind to compositional patchiness

We concentrate our study on the statistical analysis of 121 DNA sequences selected in the human genome, with the requirement that their overall lengh $L \geqslant 2000$ nucleotides, so that the range of scales available to fractal scaling be large enough to make the analysis meaningful with respect to finite size effects. We took the sequences from EMBL data bank and processed seperately 47 coding (individual exons,CDS's) and 74 noncoding (individual introns) regions. To graphically portray these sequences we follow the so-called "DNA walk" analysis [7] which requires first to convert the four letter (A,C,G,T) text into a binary sequence. This can be done, for example, on the basis of purine (A,G) vs. pyrimidine (C,T) distinction, by defining the incremental variable that associates to position i the value $\chi(i) = 1$ or -1, depending on whether the ith nucleotide of the sequence is a purine or a pyrimidine. (We refer the reader to Refs. [17,18,21] for similar analysis with the two complementary pair-base identifications). The wavelet analysis of the human desmoplaskin I CDS is shown in Fig. 1 [17,18]. The patchiness of this sequence is patent on the corresponding DNA walk ($f(x) = \sum_{i=1}^{x} \chi(i)$) in Fig. 1a: one clearly recognizes three regions of different strand bias. Fig. 1b shows the WT space–scale representation of this DNA signal when using the order 1 analyzing wavelet $\psi^{(1)}$. In Fig. 1c and Fig. 1d, two horizontal cuts

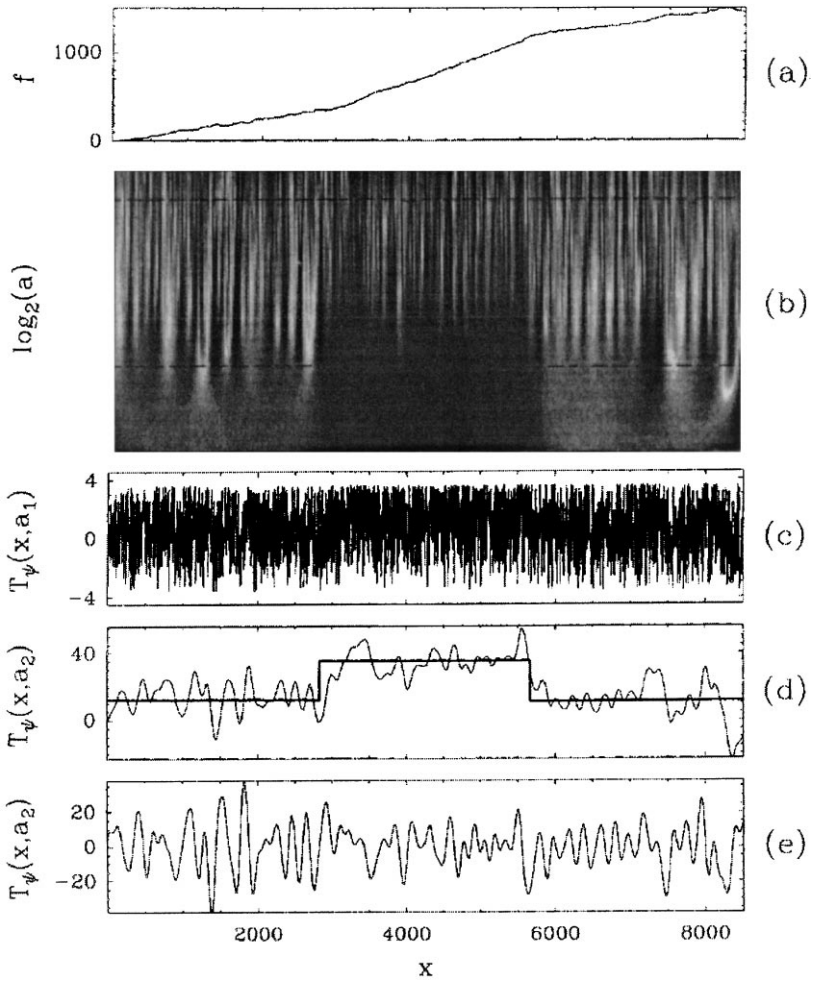

Fig. 1. WT analysis of the human desmoplakin I CDS ($L = 8499$). (a) DNA walk displacement $f(x)$ (excess of purines over pyrimidines) vs. nucleotide distance x. (b) WT of $f(x)$ computed with the analyzing wavelet $\psi^{(1)}$; $T_{\psi^{(1)}}(x,a)$ is coded, independently at each scale a, using 32 grey levels from white ($\min_x T_{\psi^{(1)}}(x,a)$) to black ($\max_x T_{\psi^{(1)}}(x,a)$); small scales are at the top. (c) $T_{\psi^{(1)}}(x, a = a_1)$ vs. x for $a_1 = 2^3$ ($\sim$32 nucleotides). (d) $T_{\psi^{(1)}}(x, a = a_2)$ vs. x for $a_2 = 2^7$ ($\sim$512 nucleotides). (e) Same analysis as in (d) but with the analyzing wavelet $\psi^{(2)}$.

$T_{\psi^{(1)}}(x,a)$ are shown at two different scales $a = a_1 = 2^3$ and $a_2 = 2^7$ which correspond (when taking into account the characteristic size of $\psi^{(1)}$) to looking at the fluctuations of the DNA walk over a characteristic lengh of the order of 32 and 512 nucleotides respectively. When progressively increasing the WT magnification, one realizes that the fluctuations detected at small scales actually occur around three successive linear trends. $\psi^{(1)}$ not being blind to linear behavior, the WT coefficients fluctuate about finite constant values that correspond to the slopes of those linear trends. This phenomenon

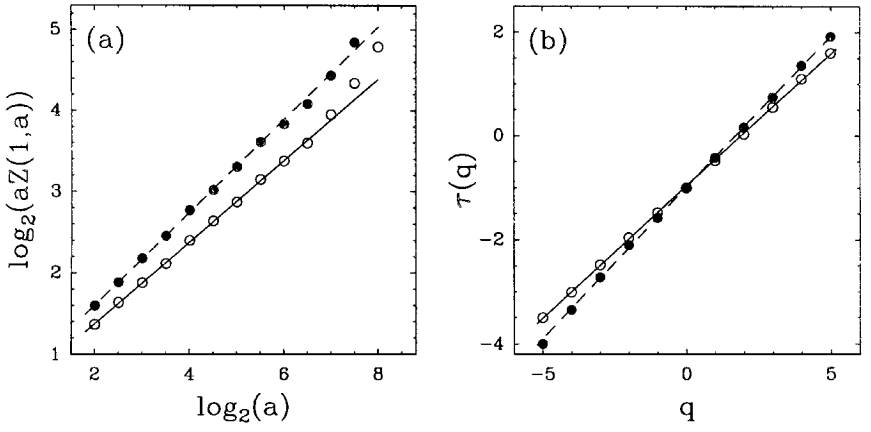

Fig. 2. Comparative WTMM analysis of the DNA walk graphs of the largest intron ($L = 71718$) of the human retinoblastoma susceptibility gene (●) and of the dystrophin CDS ($L = 11050$) (○). (a) $\log_2(aZ(q,a))$ vs. $\log_2 a$ for $q = 1$; the straight lines represent the corresponding least-square fit estimates of $1 + \tau(1) = H = 0.50 \pm 0.02$ (solid line) and 0.57 ± 0.02 (dashed line). (b) $\tau(q)$ vs. q; the solid and dashed lines correspond to the theoritical spectrum $\tau(q) = qH - 1$ for classical Brownian motion ($H = 0.50 \pm 0.01$) and fractional Brownian motion ($H = 0.58 \pm 0.02$), respectively. The analyzing wavelet is $\psi^{(2)}$.

is indeed present at all scales and drastically affects the fractal branching of the WT. In Fig. 1e, the fluctuations of the WT coefficients are shown at the same coarse-scale $a = a_2$ but as computed with the order-2 analyzing wavelet $\psi^{(2)}$. The WT microscope now being orthogonal also to linear behavior, the WT fluctuates about zero and one does not see the influence of the strand bias anymore. Furthermore, by considering successively $\psi^{(3)}$, $\psi^{(4)}, \ldots$, one can hope not only to restore the stationarity of the increments of the DNA signal but also to eliminate more complicated nonlinear trends.

3.2. Application of the WTMM method to the study of DNA sequences

In Fig. 2 are reported typical data coming out from the application of the WTMM method to the DNA walk corresponding to the dystrophin CDS when using the analyzing wavelet $\psi^{(2)}$. Fig. 2a shows plots of the partition functions $Z(q,a)$ computed from the WT skeleton according to Eq. (2) vs. the scale parameter a in a log–log representation. Only plots obtained for $q = 1$ are shown on this figure since they are quite representative of the typical features of the data for different values of q ($-2 \leqslant q \leqslant 4$). For convenience, $aZ(1,a)$ is plotted in Fig. 2a, since from Eq. (2), it is expected to scale like $a^{\tau(1)+1}$; this will allow us to compare our results for DNA walks directly to the prediction for homogeneous fBms: $aZ(1,a) \sim a^H$ where H is the Hurst exponent [17,18]. From a linear regression fit over a reasonable range of scales (more than a decade), one gets $\tau(1) + 1 = 0.50 \pm 0.02$, i.e., a value which is, up to the experimental uncertainty, quite consistent with the value $H = \frac{1}{2}$ for uncorrelated Brownian random walks. This partial result is confirmed when repeating this measurement for

different values of q; as shown in Fig. 2b, the data for the overall $\tau(q)$ spectrum remarkably fall on a straight line (the hallmark of homogeneous fractal signals) of slope $h = \partial\tau/\partial q = 0.50 \pm 0.01$. In Fig. 2 are also reported the results of a similar WTMM analysis of the largest intron ($L = 71\,718$) of the human retinoblastoma susceptibility gene. In Fig. 2b the data points for $\tau(q)$ again fall on a straight line but with a slope $H = 0.58 \pm 0.02$ which is now significantly larger than $\frac{1}{2}$. Note that again the value obtained in Fig. 2a for $\tau(1)+1 = H = 0.57$ is quite compatible with the slope of $\tau(q)$. The dashed line in Fig. 2b corresponds to the theoretical spectrum for fBm's with a Hurst exponent $H = 0.58$; the remarkable agreement observed with the WTMM data confirms the presence of long-range correlations in the considered intronic DNA walk.

The results reported in Fig. 2 are actually quite representative of the results obtained for our statistical sample of 47 coding and 74 noncoding sequences [17,18]. When averaging the partition functions over these two statistical samples, we get $\bar{Z}_C(q,a)$ and $\bar{Z}_{NC}(q,a)$ which both scale with the exponent predicted for homogeneous fBm's, i.e., $\tau(q) = qH - 1$, with a main difference which allows us to distinguish coding from noncoding sequences namely the presence of long range correlation in the latter: $\bar{H}_{NC} = 0.59 \pm 0.02$, while the former look like uncorrelated random walks: $\bar{H}_C = 0.51 \pm 0.02$. These results are illustrated in Fig. 4a where $a^{1/2}Z(1,a) \sim a^{H-1/2}$ is plotted vs. $\log_2 a$ to enlighten some possible departure of H from $\frac{1}{2}$.

3.3. About the Gaussian character of the fluctuations in DNA walk landscapes

One of the most striking result of our WTMM analysis is the fact that the $\tau(q)$ spectra extracted for all the exons and introns we have considered in the human genome, are suprisingly in remarkable agreement with the theoretical prediction for Gaussian processes. Within that prospect, we have studied the probability distribution function of wavelet coefficient values $P(T_{\psi^{(2)}}(.,a))$, as computed at a fixed scale a in the fractal scaling range [17,18]. The distribution obtained for both the coding DNA sequences of Fig. 1a and the largest intron contained in the human retinoblastoma susceptibility gene are shown in Fig. 3a and Fig. 3b, respectively. When increasing the scale parameter a, the distributions become wider, but when plotting $\ln P$ vs. $T_\psi/\sigma(a)$, where $\sigma(a)$ is the r.m.s value at scale a, all the data computed at different scales fall on the same parabola independently of the nature of the sequence. Thus, as explored through the WT microscope, the basic fluctuations in DNA walks are likely to have Gaussian statistics. The presence of long-range correlations in the human introns is, in fact, contained in the scale dependence of $\sigma(a) \sim a^H$ where $H = 0.60 \pm 0.02$ as compared to the uncorrelated random-walk value $H = 0.50 \pm 0.02$ obtained for the coding sequences.

3.4. Uncovering long-range correlations in coding DNA sequences

Because of the "period three" codon structure of coding DNA, it is natural to investigate separately the three subsequences relative to the position (1, 2 or 3) of the bases within their codons [22]. We have build up these subsequences from our

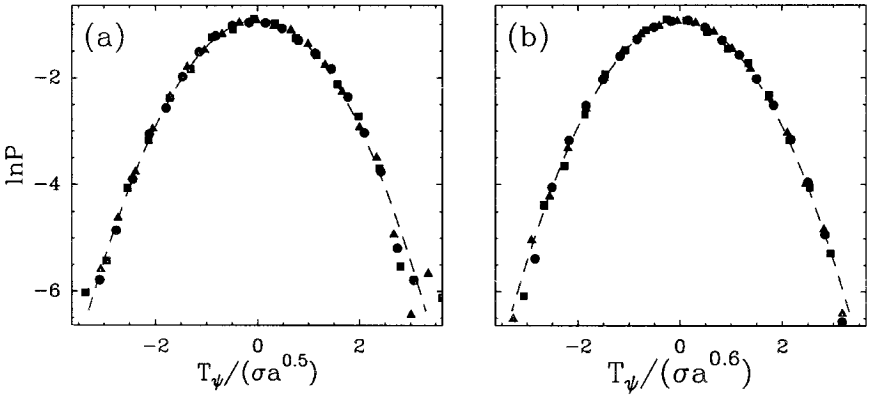

Fig. 3. Probability distribution function of wavelet coefficient values at fixed scale $a = 2^2$ (●), $a = 2^3$ (▲), $a = 2^4$ (■) corresponding approximatively to 32, 64 and 128 nucleotides; the analyzing wavelet is $\psi^{(2)}$. $\ln P$ is plotted vs. $T/\sigma(a)$, where $\sigma(a) = \sigma a^H$ is the r.m.s. value. (a) Human desmoplakin I CDS sequence: $H = 0.50$. (b) Largest intron in the human retinoblastoma susceptibility gene: $H = 0.60$. The dashed lines are parabolas characteristic of Gaussian statistics.

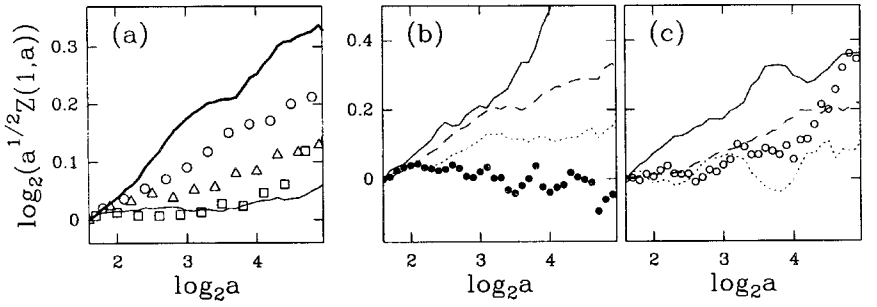

Fig. 4. $a^{1/2}Z(q = 1, a)$ vs. a in logarithmic scales, when averaged over 42 coding and 74 noncoding human DNA sequences. (a) introns (—), CDSs (—), coding subsequences relative to position 1 (△), 2 (□) and 3 (◯) of the bases whithin the codons. (b) Introns: (· · · · · ·) GC% $= 31.6 \pm 0.4$, $L = 15210$; (- - -) GC% $= 48.6 \pm 1.0$, $L = 17137$; (—) GC% $= 63.3 \pm 0.9$, $L = 10449$; (●) first intron of the human factor XIIIb subunit gene with GC% $= 31.2$, $L = 2874$. (c) Coding subsequences relative to position 3 of the bases within the codons: (· · · · · ·) GC% $= 38.1 \pm 2.9$, $L = 4759$; (- - -) GC% $= 50.8 \pm 2.8$, $L = 28521$; (—) GC% $= 62.5 \pm 2.1$, $L = 16558$; (◯) exon of the human apoB-100 gene with GC% $= 41.0$, $L = 7571$. The analyzing wavelet is $\psi^{(2)}$.

35 largest CDS sequences and we have repeated the WTMM analysis. As shown in Fig. 4a, the data for $a^{1/2}Z(1,a) \sim a^{H-1/2}$ when plotted vs. a in a log–log representation, display a rather flat behavior for both the subsequences relative to positions 1 and 2 which indicates that the corresponding roughness exponents $H_{C1} = 0.53 \pm 0.02$ and $H_{C2} = 0.51 \pm 0.02$ are undistinguishable from H_C and therefore from $\frac{1}{2}$. Surprisingly, the data for the subsequence relative to position 3 exhibit a clear linear increase with slope $\Delta H_{C3} = 0.07 \pm 0.02$ which reflects the fact that $H_{C3} = 0.57 \pm 0.02$, i.e., a value which is very close to the exponent estimated for introns. This observation suggests

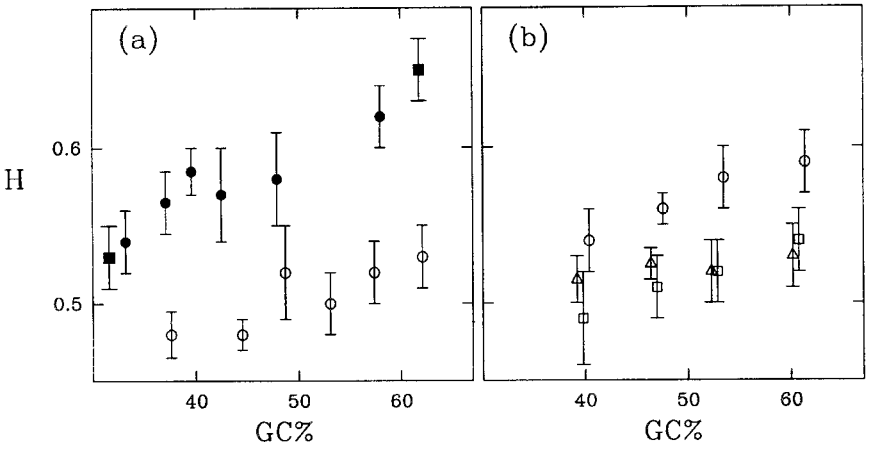

Fig. 5. WTMM estimate of the roughness exponent H vs. the GC content of the DNA sequence. (a) Introns: (●) $L \simeq 50000$, (■) $L \simeq 15000$; CDSs: (◯) $L \simeq 50000$. (b) coding subsequences relative to position 1 (△), 2 (□) and 3 (◯) of the bases within the codons: $L \simeq 20000$.

that this third coding subsequence is likely to display the same degree of long-range correlations as noncoding sequences.

3.5. Nucleotide composition effects on the long-range correlation properties of human genes

In Fig. 4b, Fig. 4c and Fig. 5 are reported the results of a similar statistical analysis when classifying these DNA sequences into categories that correspond to a given GC content [21]. The idea of looking for a link between the long-range correlation properties and the GC content of the sequences results from the remark that the WTMM method indeed fails to distinguish a few introns from actual exons [17,18]. These introns with an exponent H close to $\frac{1}{2}$, actually correspond to DNA sequences with a low GC content (from 31% to 36%). As shown in Fig. 4b, when investigating the scaling behavior of $a^{1/2}Z(1,a) \sim a^{H-1/2}$ for our set of introns, one notices some significant tendancy of the curves to become steeper when continuously increasing the GC content. The corresponding values of the roughness exponent H are reported in Fig. 5a. H clearly increases from value close to $\frac{1}{2}$ at low GC content ($\sim$30%) up to values significantly larger than 0.6 at high GC content ($>$60%). In Fig. 5a are also shown the estimates of the roughness exponent for the coding sequences. Whether the CDS be poor or rich in GC, it does not seem to possess strong long-range correlations as indicated by an exponent H close to $\frac{1}{2}$. Fig. 5b is devoted to the results of similar analysis of the 3 coding subsequences relative to the position (1, 2 or 3) of the bases within the codons. For the first and the second subsequences, one gets results quite consistent with the estimates obtained with the overall CDS sequences: whatever the GC content, the exponent H does not

significantly depart from the value $\frac{1}{2}$. Note that the data do not exclude a possible slow increase of H_{C2}. For the third subsequence, H is found to increase up to values close to 0.60 at high GC content, which brings the clue that this subsequence exhibits GC-dependent long-range correlations very much like those observed in Fig. 5a for introns (see also Fig. 4c). In order to investigate the possibility that these observations might result from the exon concatenation in the CDSs, we have analyzed individual human exons (for statistical reason, only the largest ones). These exons exhibit the same features than the CDSs, as it is exemplified in Fig. 4c by the apoB-100 largest exon.

4. Discussion

The results reported in this work clearly show that the GC content is likely to be relevant to the long-range correlation properties observed in both intronic and exonic DNA sequences. The evolution of DNA sequences in terms of GC content has attracted a lot of interest during the past few years [14–16, 22–27]. Several mechanisms can be proposed to account for the observed long-range correlations in the GC rich intronic sequences [21].

(i) Besides punctual mutations, genomic sequences are subject to a number of insertion–deletion events of DNA fragments of widely variable sizes. These events are much less frequent in exonic regions due to the strong constraints imposed by their coding properties. The insertion–deletion mechanisms could be responsible for the observed long-range correlations [2]. However, insertions–deletions occur in low GC intronic sequences which we just showed to present no long-range correlations. Furthermore, the correlations observed between the third bases of the codons, but not between adjacent nucleotides, are unlikely to result from (rare) insertion–deletion events which generally involve several adjacent nucleotides in order to maintain the coding phase.

(ii) The human genome is well known to be compartimentalized into wide specific domains with uniform GC content, called isochores [23]; appreciable scatter of the average GC content is actually observed when comparing different domains. Another hypothesis is to consider that the processes operating to create the GC-rich isochores lead to the appearence of long-range correlations. Thanks to the functional constraints acting on the coding sequences embedded in these GC-rich regions, these processes should be less active on the exons, with a concomittant lack of long-range correlations as compared to the surrounding introns. Since these constraints are less stringent on the third base of the codons, this would explain the correlations observed between these nucleotides in high GC containing exons. In human genes the frequencies of the third base of codons are highly correlated with neighboring intronic GC content [28]. This property favors the hypothesis that the exonic correlations are produced by the same mechanisms which lead to intronic correlations.

It is likely that the observations reported here also extend to genomes of mammals and warmblooded vertebrates. The exploration of genomes of various organisms including unicellular eukaryotes and prokaryotes is currently under progress.

Acknowledgements

This research was supported by the GIP GREG (project "Motifs dans les Séquences") and by the Ministère de l'Education Nationale, de l'Enseignement Supérieur, de la Recherche et de l'Insertion Professionnelle ACC-SV (project "Génétique et Environnement").

References

[1] H.E. Stanley, S.V. Buldyrev, A.L. Goldberger, S. Havlin, S.M. Ossadnik, C.K. Peng, M. Simons, Fractal 1 (1993) 283.
[2] W. Li, Int. J. Bif. Chaos 2 (1992) 137.
[3] M.Ya. Azbel, Phys. Rev. Lett. 75 (1995) 168.
[4] H. Herzel, I. Große, Physica A 216 (1995) 518.
[5] R.F. Voss, Phys. Rev. Lett. 68 (1992) 3805.
[6] R.F. Voss, Fractal 2 (1994) 1.
[7] C.K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, F. Sciortino, M. Simons, H.E. Stanley, Nature 356 (1992) 168.
[8] R.N. Mantegna, S.V. Buldyrev, A.L. Goldberger, S. Havlin, C.K. Peng, M. Simons, H.E. Stanley, Phys. Rev. E 52 (1995) 2939.
[9] S. Nee, Nature 357 (1992) 450.
[10] C.A. Chazidimitriou-Dreismann, L. Larhammar, Nature 361 (1993) 212.
[11] S. Karlin, V. Brendel, Science 259 (1993) 677.
[12] B. Borstnick, D. Pumpernik, D. Lukman, Europhys. Lett. 23 (1993) 389.
[13] S.V. Buldyrev, A.L. Goldberger, S. Havlin, R.N. Mantegna, M.E. Matsa, C.K. Peng, M. Simons, H.E. Stanley, Phys. Rev. E 51 (1995) 5084.
[14] G. Bernardi, B. Olofsson, J. Filipski, M. Zerial, J. Salinas, F. Cuny, M. Meunier-Rotival, F. Rodier, Science 228 (1985) 953.
[15] G.A. Churchill, Bull. Math. Biol. 51 (1989) 79.
[16] J.W. Fickett, D.C. Torney, D.R. Wolf, Genomics 13 (1992) 1056.
[17] A. Arneodo, E. Bacry, P.V. Graves, J.F. Muzy, Phys. Rev. Lett. 74 (1995) 3293.
[18] A. Arneodo, Y. d'Aubenton-Carafa, E. Bacry, P.V. Graves, J.F. Muzy, C. Thermes, Physica D 96 (1996) 291.
[19] J.F. Muzy, E. Bacry, A. Arneodo, Int. J. Bif. Chaos 4 (1994) 245.
[20] A. Arneodo, E. Bacry, J.F. Muzy, Physica A 213 (1995) 232.
[21] A. Arneodo, Y. d'Aubenton-Carafa, B. Audit, E. Bacry, J.F. Muzy, C. Thermes, preprint (1997) EuroPhys. J. B, in press.
[22] P. Allegrini, M. Barbi, P. Grigolini, B.J. West, Phys. Rev. E 52 (1995) 5281.
[23] G. Bernardi, Annu. Rev. Genet. 23 (1989) 637.
[24] B. Aïssani, G. D'Onofrio, D. Mouchiroud, K. Gardiner, C. Gautier, G. Bernardi, J. Mol. Evol. 32 (1991) 493.
[25] G. D'Onofrio, D. Mouchiroud, B. Aïssani, C. Gautier, G. Bernardi, J. Mol. Evol. 32 (1991) 504.
[26] G. D'Onofrio, G. Bernardi, Gene 110 (1992) 81.
[27] P. Liò, S. Ruffo, M. Buiatti, J. Theor. Biol. 171 (1994) 215.
[28] P. Sharp, G. Matassi, Curr. Opin. Gent. Dev. 4 (1994) 851.