ELSEVIER

Physica A 249 (1998) 430–438

PHYSICA A

Analysis of DNA sequences using methods of statistical physics

S.V. Buldyrev [a,*], N.V. Dokholyan [a], A.L. Goldberger [b,c], S. Havlin [a,d], C.-K. Peng [a,b], H.E. Stanley [a], G.M. Viswanathan [a]

[a] *Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215, USA*
[b] *Cardiovascular Division, Harvard Medical School, Beth Israel Hospital, Boston, MA 02215, USA*
[c] *Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA*
[d] *Gonda-Goldschmied Center and Department of Physics, Bar-Ilan University, Ramat-Gan 52900, Israel*

Abstract

We review the present status of the studies of DNA sequences using methods of statistical physics. We present evidence, based on systematic studies of the entire GenBank database, supporting the idea that the DNA sequence in genes containing *noncoding* regions is correlated, and that the correlation is remarkably long range, i.e., base pairs *thousands of base pairs* distant are correlated. We do not find such a long-range correlation in the coding regions of the DNA. We discuss the mechanisms of molecular evolution that may lead to the presence of long-range power-law correlations in noncoding DNA and their absence in coding DNA. One such mechanism is the simple repeat expansion, which recently has attracted the attention of the biological community in conjunction with genetic diseases. We also review new tools – e.g., detrended fluctuation analysis – that are useful for studies of complex hierarchical DNA structure.

PACS: 87.10.+e
Keywords: DNA; Long-range correlations; Mutations

1. Long-range power-law correlations

In recent years long-range power-law correlations have been discovered in a wide variety of systems. Such long-range power-law correlations are a physical fact that in turn gives rise to the increasingly appreciated "fractal geometry of nature" [1–17]. Recognizing the ubiquity of long-range power-law correlations can help us in our efforts to understand nature, since as soon as we find power-law correlations we can quantify them with a critical exponent. Quantification of this kind of scaling behavior

* Corresponding author.

PII S0378-4371(97)00503-7

for apparently unrelated systems allows us to recognize similarities between different systems, leading to underlying unifications that might otherwise have gone unnoticed.

Traditionally, investigators in many fields characterize processes by assuming that correlations decay exponentially. However, there is one major exception: at the critical point, the exponential decay turns into a power-law decay [18]. Many systems drive themselves spontaneously toward critical points [2–4,19].

In the following sections we will attempt to summarize some recent findings [20–49] concerning the possibility that – under suitable conditions – the sequence of base pairs or "nucleotides" in DNA also displays power-law correlations. We will also discuss the intriguing implications of this finding for molecular evolution.

The role of genomic DNA sequences in coding for protein structure is well known [50]. However, in the genomes of high eukaryotic organisms only a small portion of the total genome length is used for protein coding (as low as 3% in the human genome). The segments of the chromosomal DNA that are spliced out during the formation of a mature mRNA are called *introns* (for intervening sequences). The coding sequences are called *exons* (for expressive sequences).

The role of introns and intergenic sequences constituting large portions of the genome remains unknown. Furthermore, only a few quantitative methods are currently available for analyzing information which is possibly encrypted in the noncoding part of the genome.

2. The "DNA walk"

One interesting question that may be asked by statistical physicists would be whether the sequence of the nucleotides A, C, G and T behaves like a one-dimensional "ideal gas", where the fluctuations of density of certain particles obey Gaussian law, or if there exist long-range correlations in nucleotide content (as in the vicinity of a critical point). These result in domains of all size with different nucleotide concentrations. Such domains of various sizes were known for a long time but their origin and statistical properties remain unexplained. A natural language to describe heterogeneous DNA structure is long-range correlation analysis, borrowed from the theory of critical phenomena [18].

In order to study the scale-invariant long-range correlations of a DNA sequence, we introduce a graphical representation of DNA sequences, which we term a *fractal landscape* or *DNA walk* [20]. For the conventional one-dimensional random walk model [51,52], a walker moves either "up" [$u(i) = +1$] or "down" [$u(i) = -1$] one unit length for each step i of the walk. For the case of an uncorrelated walk, the direction of each step is independent of the previous steps. For the case of a correlated random walk, the direction of each step depends on the history ("memory") of the walker [53–60].

One definition of the DNA walk is that the walker steps "up" if a pyrimidine (C or T) occurs at position i along the DNA chain, while the walker steps "down" if a purine

(A or G) occurs at position i. The question we asked was whether such a walk displays only short-range correlations (as in an n-step Markov chain) or long-range correlations (as in critical phenomena and other scale-free "fractal" phenomena). A different kind of DNA walk was suggested by Azbel [61]. There have also been attempts to map DNA sequence onto multi-dimensional DNA walks [21,62].

The DNA walk allows one to visualize directly the fluctuations of the purine-pyrimidine content in DNA sequences: Positive slopes correspond to high concentration of pyrimidines, while negative slopes correspond to high concentration of purines. Visual observation of DNA walks suggests that the coding sequences and intron-containing noncoding sequences have quite different landscapes.

3. Correlations and fluctuations

An important statistical quantity characterizing any walk [51,52] is the root-mean-square fluctuation $F(\ell)$ about the average of the displacement of a quantity $\Delta y(\ell)$ defined by $\Delta y(\ell) \equiv y(\ell_0 + \ell) - y(\ell_0)$, where $y(\ell) \equiv \sum_{i=1}^{\ell} u(i)$. If there is no characteristic length (i.e., if the correlation between $u(i)$ and $u(j)$ are power-law long-range correlations), then fluctuations will also be described by a power law

$$F(\ell) \sim \ell^{\alpha} \tag{1}$$

with $\alpha \neq \frac{1}{2}$. The case $\alpha = \frac{1}{2}$ represents the absence of long-range correlations.

The fact that data for intron-containing and intergenic (i.e., noncoding) sequences are linear on this double logarithmic plot confirms that $F(\ell) \sim \ell^{\alpha}$. A least-squares fit produces a straight line with slope α substantially larger than $\frac{1}{2}$, thus providing direct experimental evidence for the presence of long-range correlations [20].

On the other hand, the dependence of $F(\ell)$ for coding sequences is not linear on the log–log plot: its slope undergoes a crossover from 0.5 for small ℓ to 1 for large ℓ. However, if a single patch is analyzed separately, the log–log plot of $F(\ell)$ is again a straight line with the slope close to 0.5. This suggests that within a large patch the coding sequence is almost uncorrelated. The function $F(\ell)$ was also studied for DNA sequences by Azbel [63].

4. Detrended fluctuation analysis (DFA)

The initial report [20] on long-range (scale-invariant) correlations only in noncoding DNA sequences has generated contradicting responses. Some [21,22,26,27] support our initial finding, while some [22,28,34–37,39] disagree. However, the conclusions of Refs. [23,24] and Refs. [22,28,34–37,39] are inconsistent *with one another* in that Refs. [22,39] doubt the existence of long-range correlations (even in noncoding sequences) while Refs. [23,24,28,34–37] conclude that even coding regions display long-range correlations ($\alpha > \frac{1}{2}$). Prabhu and Claverie [28] claim that their analysis of the putative

coding regions of the yeast chromosome III produces a *wide range of exponent values*, some larger than 0.5. The source of these contradicting claims may arise from the fact that, in addition to normal statistical fluctuations expected for analysis of rather short sequences, coding regions typically consist of only a few lengthy regions of alternating strand bias – and so we have nonstationarity. Hence, conventional scaling analyses cannot be applied reliably to the entire sequence but only to subsequences.

To avoid this problem, Peng et al. [45] have recently developed a method specifically adapted to handle problems associated with nonstationary sequences which they term *detrended fluctuation analysis* (DFA). The idea of the DFA method is to compute the dependence of the standard error of a linear interpolation of a DNA walk $F_d(\ell)$ on the size of the interpolation segment ℓ. The method takes into account differences in local nucleotide content and may be applied to the entire sequence which has lengthy patches. In contrast with the original $F(\ell)$ function, which has spurious crossovers even for ℓ much smaller than a typical patch size, the detrended function $F_d(\ell)$ shows linear behavior on the log–log plot for all length scales up to the characteristic patch size, which is of the order of a thousand nucleotides in the coding sequences. For ℓ close to the characteristic patch size the log–log plot of $F_d(\ell)$ has an abrupt change in its slope.

The DFA method clearly supports the difference between coding and noncoding sequences, showing that the coding sequences are less correlated than noncoding sequences for the length scales less than 1000, which is close to characteristic patch size in the coding regions. The DFA method recently has been used to identify the typical lengths of large patches composed of different nucleotide concentration (see Refs. [64,65]). These patches may represent different structural elements of 3D chromosome organization, e.g., the DNA double helix with period 10.5 bp [66] nucleosomes about 200 bp long, 30 nm fiber, looped domains of about 10^5 bp, and chromatin bands [67,68] or isochores that may consist of several million nucleotides. Such hierarchical structure of several length scales may produce effective long-range power-law correlations.

5. Systematic analysis of GenBank database

An open question in computational molecular biology is whether long-range correlations are present in both coding and noncoding DNA or only in the latter. To answer this question, Buldyrev et al. [49] recently analyzed all 33 301 coding and all 29 453 noncoding eukaryotic sequences – each of length larger than 512 base pairs (bp) – in the present release of the GenBank to determine whether there is any statistically significant distinction in their long-range correlation properties.

They find that standard fast Fourier transform (FFT) analysis indicates that *coding* sequences have practically no correlations in the range from 10 to 100 bp (spectral exponent $\beta \pm 2SD = 0.00 \pm 0.04$). Here β is defined through the relation $S(f) \sim 1/f^{\beta}$, where $S(f)$ is the Fourier transform of the correlation function, and β is related to the

long-range correlation exponent α by $\beta = 2\alpha - 1$ so that $\alpha = \frac{1}{2}$ corresponds to $\beta = 0$ (white noise).

In contrast, for *noncoding* sequences, the average value of the spectral exponent β is positive (0.16 ± 0.05), which unambiguously shows the presence of long-range correlations. They also separately analyzed the 874 coding and 1157 noncoding sequences which have more than 4096 bp, and found a larger region of power-law behavior. They calculated the probability that these two data sets (coding and noncoding) were drawn from the same distribution, and found that it is less than 10^{-10}. They also obtained independent confirmation of these findings using the DFA method, which is designed to treat sequences with statistical heterogeneity such as DNA's known mosaic structure ("patchiness") arising from nonstationarity of nucleotide concentration. The near-perfect agreement between the two independent analysis methods, FFT and DFA, increases the confidence in the reliability of the conclusion that long-range correlation properties of coding and noncoding sequences.

From a practical viewpoint, the statistically significant difference in long-range power-law correlations between coding and noncoding DNA regions supports the development of gene-finding algorithms based on these distinct scaling properties. A recently reported algorithm of this kind [46–48] is especially useful in the analysis of DNA sequences with relatively long coding regions, such as those in yeast chromosome III.

Recently, Arneodo et al. [69] studied long-range correlation in DNA sequences using wavelet analysis. The wavelet transform can be made blind to "patchiness" of genomic sequences. They found the existence of strong long-range correlations in noncoding regimes, and weaker long-range correlations in coding regimes in excellent agreement with [49]. More recently, they found that the correlations in coding DNA are restricted to the third degenerate nucleotide of the codon [67], which may be related to the presence of the isochores, i.e., long regions of high or low C+G content [68].

6. Possible origin of long-range correlations

Long-range correlations of different length scales may develop due to different mutational mechanisms. The longest correlations, on the length scales of isochores may originate due to base-substitution mutations during replication (see Ref. [68]). Indeed, it is known that different parts of chromosomes replicate at different stages of cell division. The regions rich in C+G replicate earlier than those rich in A+T. On the other hand, the concentration of C+G precursors in the cell depletes during replication. Thus, the probability of substituting A/T for C/G is higher in those parts of the chromosome that replicate earlier. These unequal mutation rates may lead to the formation of isochores [68]. Correlations on the intermediate length scale of thousands of bp may originate due to DNA shuffling by insertion or deletion [42,44] of transposable elements such as LINES and SINES [70,71] or due to a mutation-duplication process proposed by W. Li [72] (see also Ref. [73]).

Finally, the correlations on the length scale of several hundreds of bp may evolve due to simple-repeat expansion [74]. The distributions of simple repeats are dramatically different in coding and noncoding DNA. In coding DNA they have an exponential distribution; in noncoding DNA they have long tails that in many cases may be fit by a power-law function.

The power-law distribution of simple repeats can be explained if one assumes a random multiplicative process for the mutation of the repeat length, i.e., each mutation leads to a change of repeat length by a random factor with a certain distribution (see Ref. [74]). Such a process may take place due to errors in replication [75] or unequal crossing over (see Ref. [74] and references therein). Simple-repeat expansion in the coding regions would lead to a loss of protein functionality (as, e.g., in Huntington's disease [75]) and to the extinction of the organism.

Thus, the weakness of long-range correlations in coding DNA is probably related to the coding DNA's conservation during biological evolution. Indeed, the proteins of bacteria and humans has many common templates, while the noncoding regions can be totally different even for closely related species. The conservation of protein coding sequences and the weakness of correlations in the amino acid sequences [76] are probably related to the problem of protein folding. Monte-Carlo simulations of protein folding on the cubic lattice suggest that the statistical properties of the sequences that fold into a native state resemble those of random sequences [77].

The higher tolerance of noncoding regions to various mutations, especially to mutations involving the growth of DNA length – e.g., duplication, insertion of transposable elements, and simple repeat expansion – lead to strong long-range correlations in the noncoding DNA. Such tolerance is a necessary condition for biological evolution, since its main pathway is believed to be gene duplication by chromosomal rearrangements, which does not affect coding regions [78]. However, the payoff for this tolerance is the growth of highly correlated junk DNA.

Acknowledgements

We are grateful to many individuals, including R. Mantegna, M.E. Matsa, S.M. Ossadnik, F. Sciortino and M. Simons for major contributions to those results reviewed here that represent collaborative research efforts. We also wish to thank C. Cantor, C. DeLisi, M. Frank-Kamenetskii, A.Yu. Grosberg, I. Labat, L. Liebovitch, G.S. Michaels, P. Munson, R. Nussinov, R.D. Rosenberg, E.I. Shakhnovich, M.F. Shlesinger and E.N. Trifonov for valuable discussions. Partial support was provided by the National Science Foundation, National Institutes of Health (Human Genome Project), the G. Harold and Leila Y. Mathers Charitable Foundation, the Israel–USA Binational Science Foundation, Israel Academy of Sciences, and (to C.-K.P.) by an NIH/NIMH First Award.

References

[1] B.B. Mandelbrot, The Fractal Geometry of Nature, Freeman, San Francisco, 1982.
[2] A. Bunde, S. Havlin (Eds.), Fractals and Disordered Systems, Springer, Berlin, 1991.
[3] A. Bunde, S. Havlin (Eds.), Fractals in Science, Springer, Berlin, 1994.
[4] T. Vicsek, M. Shlesinger, M. Matsushita (Eds.), Fractals in Natural Sciences, World Scientific, Singapore, 1994.
[5] J.M. Garcia-Ruiz, E. Louis, P. Meakin, L. Sander (Eds.), Growth Patterns in Physical Sciences and Biology, Proc. 1991 NATO Advanced Research Workshop, Granada, Spain, October 1991, Plenum, New York, 1993.
[6] A.Yu. Grosberg, A.R. Khokhlov, Statistical Physics of Macromolecules (Y.A. Atanov, Trans.) AIP Press, New York, 1994.
[7] J.B. Bassingthwaighte, L.S. Liebovitch, B.J. West, Fractal Physiology, Oxford University Press, New York, 1994.
[8] A.-L. Barabási, H.E. Stanley, Fractal Concepts in Surface Growth, Cambridge University Press, Cambridge, 1995.
[9] B.J. West, M.F. Shlesinger, Am. Sci. 78 (1990) 40.
[10] B.J. West, Fractal Physiology and Chaos in Medicine, World Scientific, Singapore, 1990.
[11] B.J. West, W. Deering, Phys. Rep. 246 (1994) 1.
[12] S.V. Buldyrev, A.L. Goldberger, S. Havlin, C.-K. Peng, H.E. Stanley, in: A. Bunde, S. Havlin (Eds.), Fractals in Science, Springer, Berlin, 1994, pp. 49–83.
[13] T. Vicsek, Fractal Growth Phenomena, 2nd ed., World Scientific, Singapore, 1992.
[14] J. Feder, Fractals, Plenum, New York, 1988.
[15] D. Stauffer, H.E. Stanley, From Newton to Mandelbrot: A Primer in Theoretical Physics, Springer, Heidelberg, 1990.
[16] E. Guyon, H.E. Stanley, Les Formes Fractales, Palais de la Découverte, Paris, 1991 (English Trans.: Fractal Forms, Elsevier, North-Holland, Amsterdam, 1991).
[17] H.E. Stanley, N. Ostrowsky (Eds.), Random Fluctuations and Pattern Growth: Experiments and Models, Proc. 1988 Cargèse NATO ASI, Kluwer Academic Publishers, Dordrecht, 1988.
[18] H.E. Stanley, Introduction to Phase Transitions and Critical Phenomena, Oxford University Press, London, 1971.
[19] H.E. Stanley, N. Ostrowsky (Eds.), Correlations and Connectivity: Geometric Aspects of Physics, Chemistry and Biology, Proc. 1990 Cargèse Nato ASI, Series E: Applied Sciences, Kluwer, Dordrecht 1990.
[20] C.-K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, F. Sciortino, M. Simons, H.E. Stanley, Nature 356 (1992) 168.
[21] W. Li, K. Kaneko, Europhys. Lett. 17 (1992) 655.
[22] S. Nee, Nature 357 (1992) 450.
[23] R. Voss, Phys. Rev. Lett. 68 (1992) 3805.
[24] R. Voss, Fractals 2 (1994) 1.
[25] J. Maddox, Nature 358 (1992) 103.
[26] P.J. Munson, R.C. Taylor, G.S. Michaels, Nature 360 (1992) 636.
[27] I. Amato, Science 257 (1992) 747.
[28] V.V. Prabhu, J.-M. Claverie, Nature 357 (1992) 782.
[29] P. Yam, Sci. Am. 267(3) (1992) 23.
[30] C.-K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, F. Sciortino, M. Simons, H.E. Stanley, Physica A 191 (1992) 25.
[31] H.E. Stanley, S.V. Buldyrev, A.L. Goldberger, J.M. Hausdorff, S. Havlin, J. Mietus, C.-K. Peng, F. Sciortino, M. Simons, Physica A 191 (1992) 1.
[32] H.E. Stanley, S.V. Buldyrev, A.L. Goldberger, S. Havlin, R.N. Mantegna, S.M. Ossadnik, C.-K. Peng, F. Sciortino, M. Simons, Fractals in biology and medicine, in: A. Pekalski (Ed.), Diffusion Processes: Experiment, Theory, Simulations, Proc. 5th M. Born Symp., Springer, Berlin, 1994, pp. 147–178.
[33] H.E. Stanley, S.V. Buldyrev, A.L. Goldberger, Z.D. Goldberger, S. Havlin, R.N. Mantegna, S.M. Ossadnik, C.-K. Peng, M. Simons, Statistical mechanics in biology: how Ubiquitous are long-range correlations? Proc. Int. Conf. on Statistical Mechanics, Physica A 205 (1994) 214.

[34] C.A. Chatzidimitriou-Dreismann, D. Larhammar, Nature 361 (1993) 212.
[35] D. Larhammar, C.A. Chatzidimitriou-Dreismann, Nucleic Acids Res. 21 (1993) 5167.
[36] C.A. Chatzidimitriou-Dreismann, R.M.F. Streffer, D. Larhammar, Biochim. Biophys. Acta. 1217 (1994) 181.
[37] C.A. Chatzidimitriou-Dreismann, R.M.F. Streffer, D. Larhammar, Eur. J. Biochem. 224 (1994) 365.
[38] A.Yu. Grosberg, Y. Rabin, S. Havlin, A. Neer, Europhys. Lett. 23 (1993) 373.
[39] S. Karlin, V. Brendel, Science 259 (1993) 677.
[40] C.-K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, M. Simons, H.E. Stanley, Phys. Rev. E 47 (1993) 3730.
[41] N. Shnerb, E. Eisenberg, Phys. Rev. E 49 (1994) R1005.
[42] S.V. Buldyrev, A.L. Goldberger, S. Havlin, C.-K. Peng, M. Simons, H.E. Stanley, Phys. Rev. E 47 (1993) 4514.
[43] A.S. Borovik, A.Yu. Grosberg, M.D. Frank Kamenetski, J. Biomolec. Struct. Dyn. 12 (1994) 655.
[44] S.V. Buldyrev, A.L. Goldberger, S. Havlin, C.-K. Peng, H.E. Stanley, M.H.R. Stanley, M. Simons, Biophys. J. 65 (1993) 2673.
[45] C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, Phys. Rev. E 49 (1994) 1685.
[46] S.M. Ossadnik, S.V. Buldyrev, A.L. Goldberger, S. Havlin, R.N. Mantegna, C.-K. Peng, M. Simons, H.E. Stanley, Biophys. J. 67 (1994) 64.
[47] H.E. Stanley, S.V. Buldyrev, A.L. Goldberger, S. Havlin, C.-K.Peng, M. Simons, [Proc. Int. Conf. on Condensed Matter Physics, Bar-Ilan] Physica A 200 (1993) 4.
[48] H.E. Stanley, S.V. Buldyrev, A.L. Goldberger, S. Havlin, S.M. Ossadnik, C.-K. Peng, M. Simons, Fractals 1 (1993) 283–301.
[49] 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.
[50] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, J.D. Watson, Molecular Biology of the Cell, 3rd ed., Garland Publishing, New York, 1994.
[51] E.W. Montroll, M.F. Shlesinger, The wonderful world of random walks, in: J.L. Lebowitz, E.W. Montroll (Eds.), Nonequilibrium Phenomena II. From Stochastics to Hydrodynamics, North-Holland, Amsterdam, 1984, pp. 1–121.
[52] G.H. Weiss, Random Walks, North-Holland, Amsterdam, 1994.
[53] S. Havlin, R. Selinger, M. Schwartz, H.E. Stanley, A. Bunde, Phys. Rev. Lett. 61 (1988) 1438.
[54] S. Havlin, M. Schwartz, R. Blumberg Selinger, A. Bunde, H.E. Stanley, Phys. Rev. A 40 (1989) 1717.
[55] R.B. Selinger, S. Havlin, F. Leyvraz, M. Schwartz, H.E. Stanley, Phys. Rev. A 40 (1989) 6755.
[56] C.-K. Peng, S. Havlin, M. Schwartz, H.E. Stanley, G.H. Weiss, Physica A 178 (1991) 401.
[57] C.-K. Peng, S. Havlin, M. Schwartz, H.E. Stanley, Phys. Rev. A 44 (1991) 2239.
[58] M. Araujo, S. Havlin, G.H. Weiss, H.E. Stanley, Phys. Rev. A 43 (1991) 5207.
[59] S. Havlin, S.V. Buldyrev, H.E. Stanley, G.H. Weiss, J. Phys. A 24 (1991) L925.
[60] S. Prakash, S. Havlin, M. Schwartz, H.E. Stanley, Phys. Rev. A 46 (1992) R1724.
[61] M.Y. Azbel, Phys. Rev. Lett. 31 (1973) 589.
[62] C.L. Berthelsen, J.A. Glazier, M.H. Skolnick, Phys. Rev. A 45 (1992) 8902.
[63] M.Y. Azbel, Biopolymers 21 (1982) 1687.
[64] G.M. Viswanathan, S.V. Buldyrev, S. Havlin, H.E. Stanley, Biophys. J. 72 (1997) 866–875.
[65] G.M. Viswanathan, S.V. Buldyrev, S. Havlin, H.E. Stanley, Physica A 249 (1998) 581–586, these Proceedings.
[66] E.N. Trifonov, Physica A 249 (1998) 511–516, these Proceedings; H. Herzel, E.N. Trifonov, O. Weiss, I. Große, Physica A 249 (1998) 449–459, these Proceedings.
[67] A. Arneodo, Y. D'Aubenton-Carafa, B. Audit, E. Bacry, J.F. Muzy, C. Thermes, Physica A 249 (1998) 439–448, these Proceedings.
[68] X. Gu, W.-H. Li, J. Mol. Evol. 38 (1994) 468–475.
[69] A. Arneodo, E. Bacry, P.V. Graves, J.F. Muzy, Phys. Rev. Lett. 74 (1995) 3293–3296.
[70] J. Jurka, T. Walichiewicz, A. Milosavljevic, J. Mol. Evol. 35 (1992) 286.
[71] R.H. Hwu, J.W. Roberts, E.H. Davidson, R.J. Britten, Proc. Nat. Acad. Sci. USA. 83 (1986) 3875.
[72] W.-H. Li, Phys. Rev. A 43 (1991) 5240–5260.

[73] H.E. Stanley, V. Afanasyev, L.A.N. Amaral, S.V. Buldyrev, A.L. Goldberger, S. Havlin, H. Leschhorn, P. Maass, R.N. Mantegna, C.-K. Peng, P.A. Prince, M.A. Salinger, M.H.R. Stanley, G.M. Viswanathan, Physica A 224 (1996) 302–321.
[74] N.V. Dokholyan, S.V. Buldyrev, S. Havlin, H.E. Stanley, Physica A 249 (1998) 594–599, these Proceedings; Phys. Rev. Lett., in press.
[75] R.D. Wells, J. Biol. Chem. 271 (1996) 2875–2878.
[76] V. Pande, A.Ya. Gosberg, T. Tanaka, Proc. Nat. Acad. Sci. USA 91 (1994) 12 972.
[77] E.I. Shakhnovich, A.M. Gutin, Nature 346 (1990) 773.
[78] W.-H. Li, T.G. Marr, K. Kaneko, Physica D 75 (1994) 392–416.