Pergamon

Chaos, Solitons & Fractals Vol. 6, pp. 171–201, 1995
Elsevier Science Ltd
Printed in Great Britain
0960-0779/95 $9.50 + .00

0960-0779(94)00259-2

Fractals in Biology and Medicine

S. HAVLIN,*‡ S.V. BULDYREV,* A.L. GOLDBERGER,† R.N. MANTEGNA,* S.M. OSSADNIK,* C.-K. PENG,†* M. SIMONS†§ and H.E. STANLEY*

* *Center for Polymer Studies and Department of Physics, Boston University, Boston, MA USA*

‡ *Department of Physics, Bar-Ilan University, Ramat-Gan, Israel*

† *Cardiovascular Division, Harvard Medical School, Beth Israel Hospital, Boston, MA 02215 USA*

§ *Department of Biology, MIT, Cambridge, MA 02139 USA*

Abstract – Our purpose is to describe some recent progress in applying fractal concepts to systems of relevance to biology and medicine. We review several biological systems characterized by fractal geometry, with a particular focus on the long-range power-law correlations found recently in DNA sequences containing *noncoding* material. Furthermore, we discuss the finding that the exponent α quantifying these long-range correlations ("fractal complexity") is smaller for coding than for noncoding sequences. We also discuss the application of fractal scaling analysis to the dynamics of heartbeat regulation, and report the recent finding that the normal heart is characterized by long-range "anticorrelations" which are absent in the diseased heart.

1 Introduction

In the last decade it was realized that some biological systems have no characteristic length or time scale, i.e., they have fractal—or, more generally, self-affine—properties [1,2]. However, the fractal properties in different biological systems, have quite different nature, origin, and appearance. In some cases, it is the geometrical shape of a biological object itself that exhibits obvious fractal features, while in other cases the fractal properties are more "hidden" and can only be perceived if data are studied as a function of time or mapped onto a graph in some special way. After an appropriate mapping, such a graph may resemble a mountain landscape, with jagged ridges of all length scales from very small bumps to enormous peaks. Mathematically, these landscapes can be quantified in terms of fractal concepts such as self-affinity. The main part of the chapter is devoted to the study of such hidden fractal properties that have been recently discovered in DNA sequences and heartbeat activity.

2 Fractal Shapes

In contrast to compact objects, fractal objects have a very large *surface* area. In fact, they are composed almost entirely of "surface." This observation explains why fractals are ubiquitous in biology, where surface phenomena

SSDI

are of crucial importance.

Lungs exemplify this feature (Fig. 1). The surface area of a human lung is as large as a tennis court. The mammalian lung is made up of self-similar branches with many length scales, which is the defining attribute of a fractal surface. The efficiency of the lung is enhanced by this fractal property, since with each breath oxygen and carbon dioxide have to be exchanged at the lung surface. The structure of the bronchial tree has been quantitatively analyzed using fractal concepts [2,4]. In particular, fractal geometry could explain the power law decay of the average diameter of the bronchial tube with the generation number, in contrast to the classical model which predicts an exponential decay [6].

Not only the geometry of the respiratory tree is described by fractal geometry, but also the time-dependent features of inspiration. Specifically, Suki et al. [5] studied airway opening in isolated dog lungs. During constant flow inflations, they found that the lung volume changes in discrete jumps (Fig. 1), and that the probability distribution function of the relative size x of the jumps, $\Pi(x)$, and that of the time intervals t between these jumps, $\Pi(t)$, follow a power law over nearly two decades of x and t with exponents of 1.8 and 2.7, respectively. To interpret these findings, they developed a branching airway model in which airways, labeled ij, are closed with a uniform distribution of opening threshold pressures P. When the "airway opening" pressure P_{ao} exceeds P_{ij} of an airway, that airway opens along with one or both of its daughter branches if $P_{ij} < P_{ao}$ for the daughters. Thus, the model predicts "avalanches" of airway openings with a wide distribution of sizes, and the statistics of the jumps agree with those $\Pi(x)$ and $\Pi(t)$ measured experimentally. They concluded that power law distributions, arising from avalanches triggered by threshold phenomena, govern the recruitment of terminal airspaces.

A second example is the arterial system which delivers oxygen and nutrients to all the cells of the body. For this purpose blood vessels must have fractal properties [7,8]. The diameter distribution of blood vessels ranging from capillaries to arteries follows a power-law distribution which is one of the main characteristics of fractals. Sernetz et al. [9] have studied the branching patterns of arterial kidney vessels. They analyzed the mass-radius relation and found that it can be characterized by fractal geometry, with fractal dimensions between 2.0 and 2.5. Similarly, the branching of trees and other plants, as well as root systems have a fractal nature [10]. Moreover, the size distribution of plant-supported insects was found to be related to the fractal distribution of the leaves [11].

One of the most remarkable examples of a fractal object is the surface of a cauliflower, where every little head is an "almost" exact reduced copy of the whole head formed by intersecting Fibonacci spirals of smaller heads, which in turn consist of spirals of smaller and smaller heads, up to the fifth order of hierarchy (see Fig. 8.0 in [3]). West and Goldberger were first to describe such a "Fibonacci fractal" in the human lung [2].

Considerable interest in the biological community has also arisen from the possibility that neuron shape can be quantified using fractal concepts. For example, Smith et al. [12] studied the fractal features of vertebrate central nervous system neurons in culture and found that the fractal dimension is increased as the neuron becomes more developed. Caserta et al. [13] showed that the shapes of quasi-two-dimensional retinal neurons can be characterized by a fractal dimension d_f. They found for fully developed neurons *in vivo*, $d_f = 1.68 \pm 0.15$, and suggest that the growth mechanism for neurite outgrowth bears a direct analogy with the growth model called *diffusion limited aggregation* (DLA). The branching pattern of retinal vessels in a developed human eye is also similar to DLA [8]. The fractal dimension was estimated to be about 1.7, in good agreement with DLA for the case of two dimensions. For an alternative model for retinal growth see [14].

The DLA-type model governing viscous fingering may also serve to resolve the age-old paradox *"Why doesn't the stomach digest itself?* [15]. Indeed, the concentration of hydrochloric acid in the mammalian stomach after each meal is sufficient to digest the stomach itself, yet the gastric epithelium normally remains undamaged in this harsh environment. One protective factor is gastric mucus, a viscous secretion of specialized cells, which forms a protective layer and acts as a diffusion barrier to acid. Bicarbonate ion secreted by the gastric epithelium is trapped in the mucus gel, establishing a gradient from pH 1-2 at the lumen to pH 6-7 at the cell surface. The

Fig. 1. The dynamic mechanism responsible for filling the lung involves "avalanches" or "bursts" of air that occur in all sizes—instead of an exponential distribution, one finds a power law distribution [5]. The underlying cause of this scale-free distribution of avalanches is the fact that every airway in the lung has its own threshold below which it is not inflated. Shown here is a diagram of the development of avalanches in the airways during airway opening. At first, almost all airways whose threshold value is smaller than the external pressure (red) are closed. Then the airway opening pressure increases until a second threshold is exceeded, and as a result all airways further up the tree whose thresholds are smaller become inflated (green). The airway opening pressure is successively increased until third, fourth, and fifth thresholds are exceeded (yellow, brown, and blue). The last threshold to be exceeded results in filling the airways colored violet; we notice that this last avalanche opens up over 25% of the total lung volume, thereby significantly increasing the total surface area available for gas exchange. After [5]

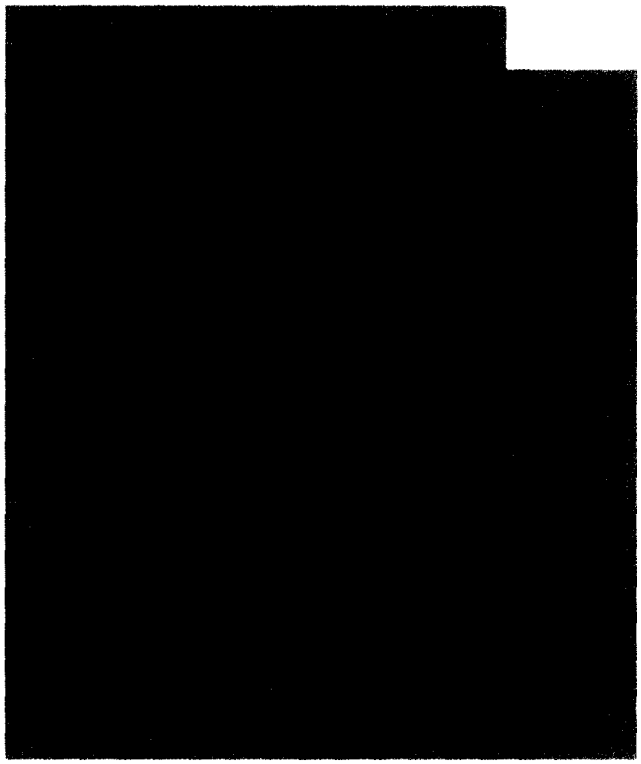

Fig. 2. Photograph of a retinal neuron (nerve cell), the morphology is similar to the DLA archetype. After [13]

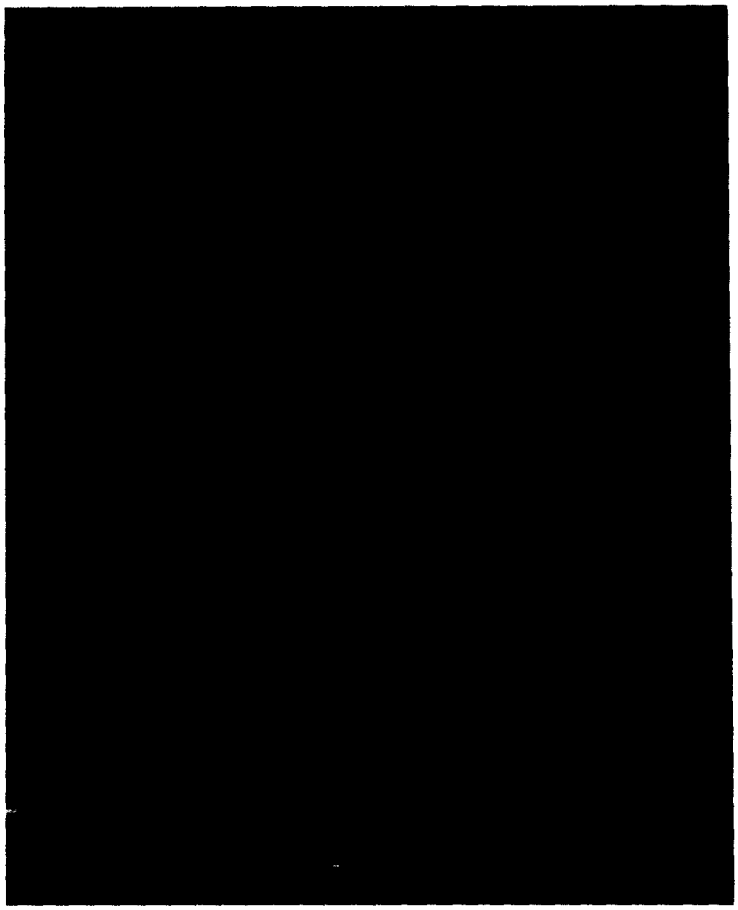

Fig. 3. Viscous fingers reflect the complex interface that develops when one fluid is pumped through another of higher viscosity. Shown is the formation of such viscous fingers or channels when hydrochloric acid is injected into solutions of gastric mucin. These channels may confine the acid and direct it to the lumen, thus protecting the gastric mucosa from acidification and ulceration; when the gastric glands contract, acid is ejected under high enough pressure to form viscous fingers. After [15]

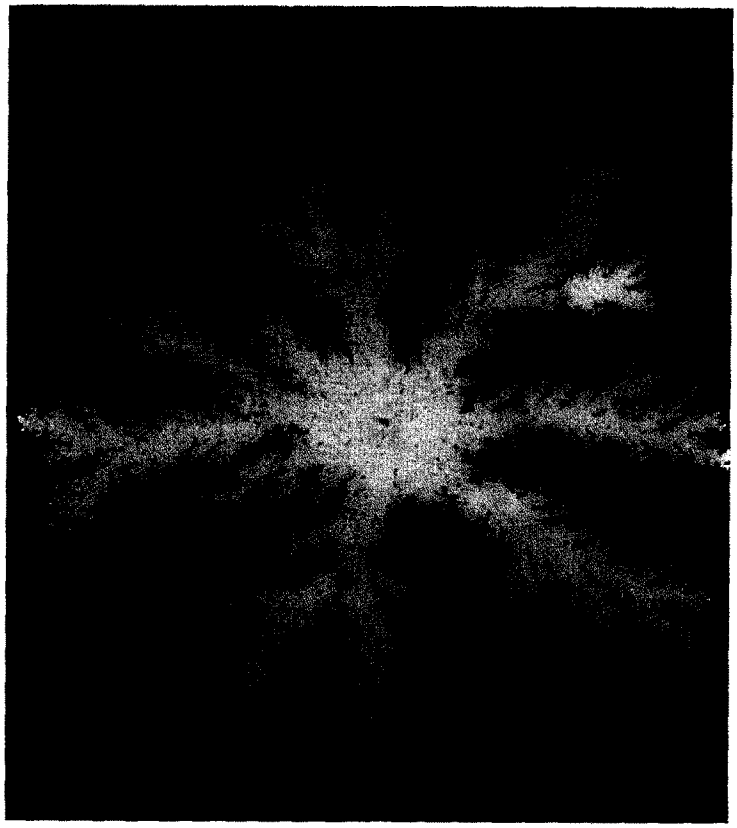

Fig. 4. A typical example of DLA-like colony patterns incubated at 35^{O}C for three weeks after inoculation on the surface of agar plates containing initially 1 g/ℓ of peptone as nutrient. This pattern has a fractal dimension of $\{d_f \cong 1.72\}2$. After Matsushita and Fujikawa [16]

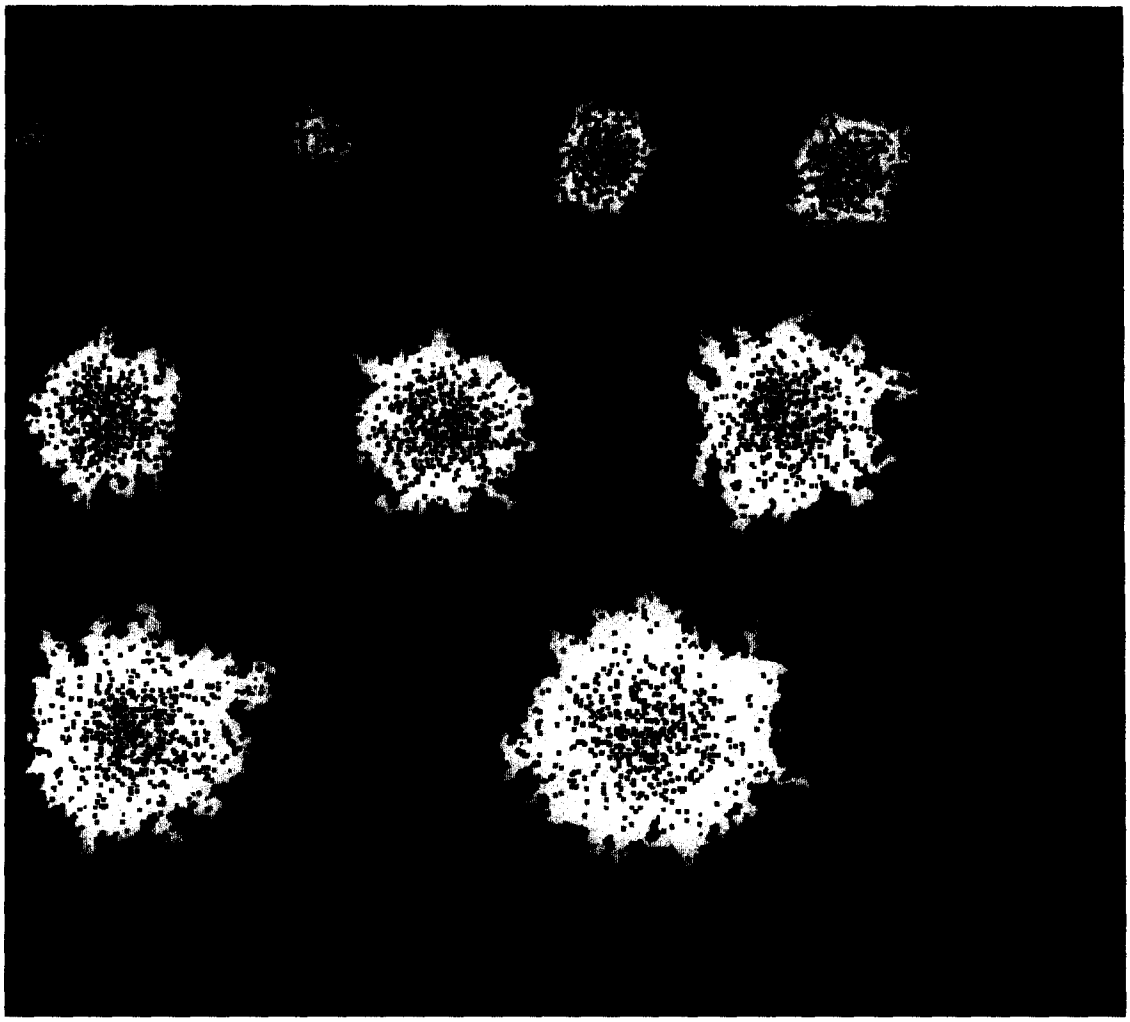

Fig. 5. Snapshots at successive times of the territory covered by N random walkers for the case $N = 500$ for a sequence of times. Note the roughening of the disc surface as time increases. The roughening is characteristic of the experimental findings for the diffusive spread of a population [21]. After [20], courtesy of P. Trunfio

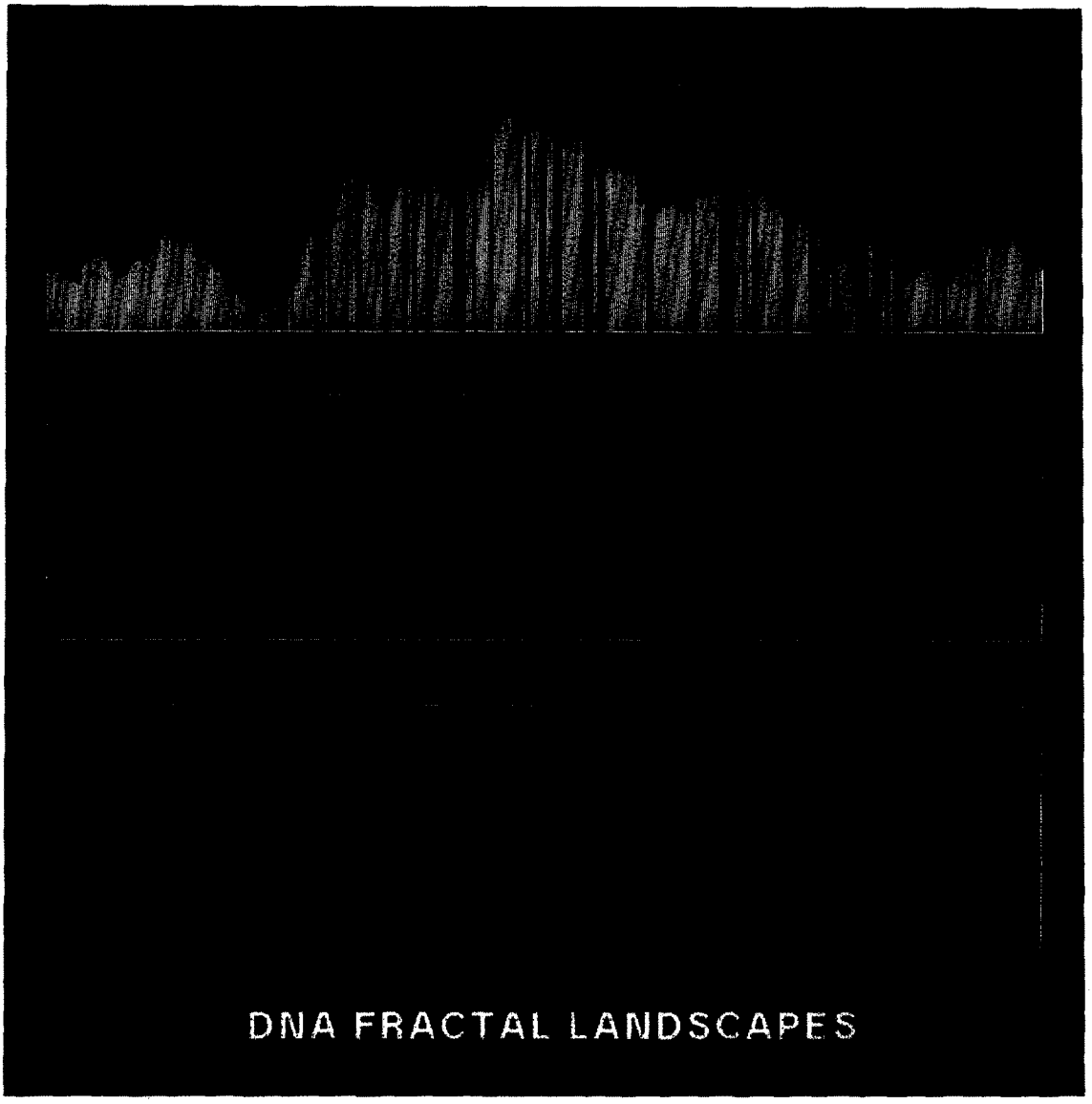

Fig. 8. The DNA walk representations of (a) human β-cardiac myosin heavy chain gene sequence, showing the coding regions as vertical golden bars, (b) the spliced together coding regions, and (c) the bacteriophage lambda DNA which contains only coding regions. Note the more complex fluctuations for (a) compared with the coding sequences (b) and (c). It is found that for almost all coding sequences studied that there appear regions with one strand bias, followed by regions of a different strand bias. In this presentation different step heights for purine and pyrimidine are used in order to align the end point with the starting point. This procedure is for graphical display purposes only (to allow one to visualize the fluctuations more easily) and is not used in any analytic calculations. After [33]

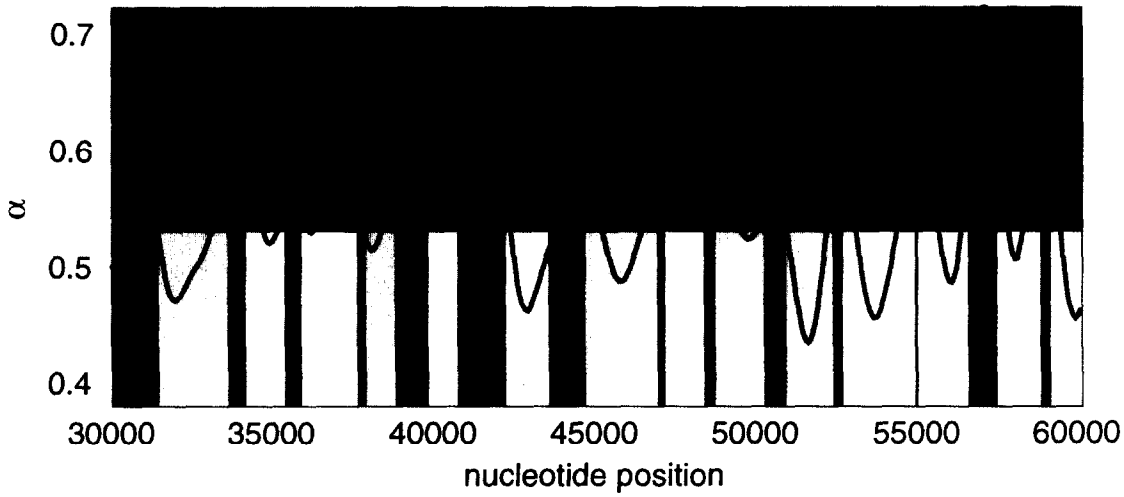

Fig. 10. Analysis of section of Yeast Chromosome III using the sliding box *Coding Sequence Finder* "CSF" algorithm. The value of the long-range correlation exponent α is shown as a function of position along the DNA chain. In this figure, the results for about 10% of the DNA are shown (from base pair #30,000 to base pair #60,000). Shown as vertical bars are the putative genes and open reading frames; denoted by the letter "G" are those genes that have been more firmly identified (March 1993 version of *GenBank*). Note that the local value of α displays **minima** where genes are suspected, while between the genes α displays **maxima**. This behavior corresponds to the fact that the DNA sequence of genes lacks long-range correlations ($\alpha = 0.5$ in the idealized limit), while the DNA sequence in between genes possesses long-range correlations ($\alpha \approx 0.6$). After [57]

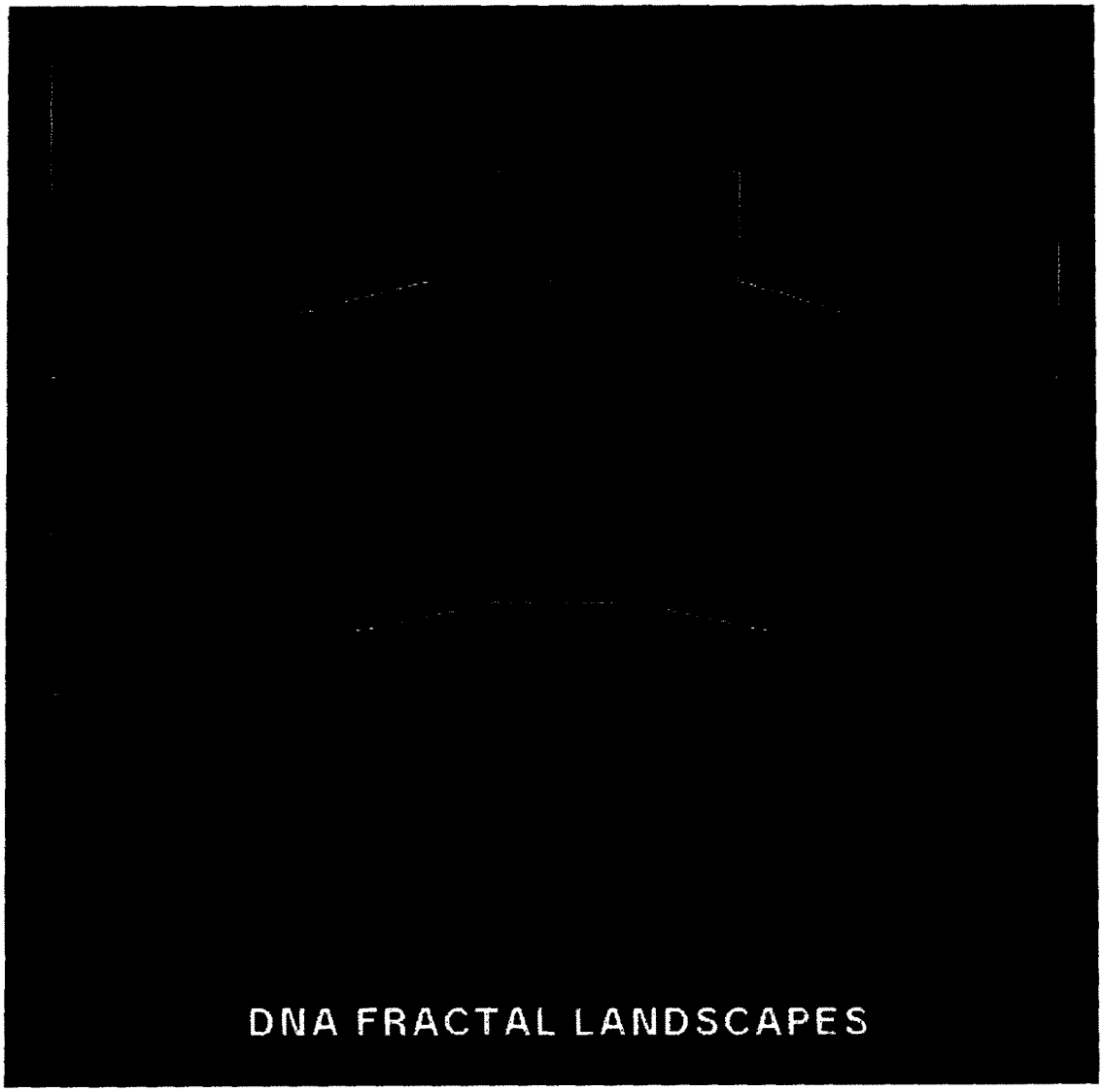

Fig. 11. The DNA walk representation for the rat embryonic skeletal myosin heavy chain gene ($\alpha = 0.63$). At the top the entire sequence is shown. In the middle the solid box shown in the top is magnified. At the bottom the solid box shown in the middle is magnified. The statistical self-similarity of these plots is consistent with the existence of a scale-free or fractal phenomenon which we call a fractal landscape. Note that one must magnify the segment by different factors along the ℓ (horizontal) direction and the y (vertical) direction; since F has the same units (dimension) as y, these magnification factors M_ℓ and M_y (along ℓ and y directions respectively) are related to the scaling exponent α by the simple relation $\alpha = \log(M_y)/\log(M_\ell)$ [e.g., from top to middle, $\log(M_y)/\log(M_\ell) = \log(2.07)/\log(3.2) = 0.63$]

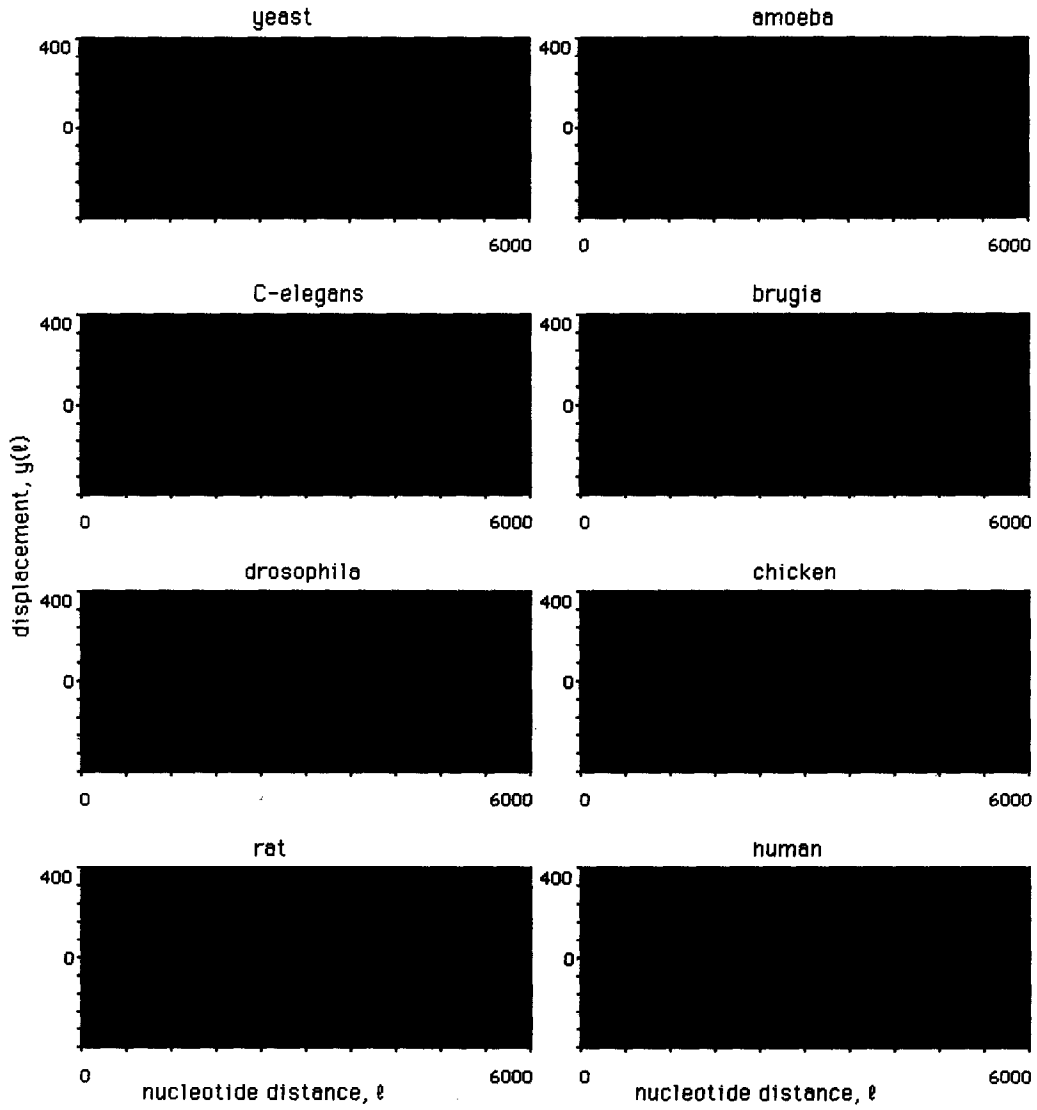

Fig. 12. The DNA walk representations of (a) 8 cDNA sequences from the MHC family and (b) the corresponding genes. DNA landscapes are plotted so that the end points have the same vertical displacement as the starting points [33]. The graphs are for yeast, amoeba, worms: *C. elegans*, *Brugia malayi*, drosophila, chicken, rat and human (from top to bottom, left to right). The shaded areas in (b) denote coding regions of the genes. The DNA walks for the genes show increasing "complexity" with evolution. In contrast, the cDNA walks all show remarkably similar crossover patterns due to sequential "up-hill" and "down-hill" slopes representing different purine/pyrimidine strand biases in the regions coding for the head and tail of the MHC molecule, respectively. After [55]

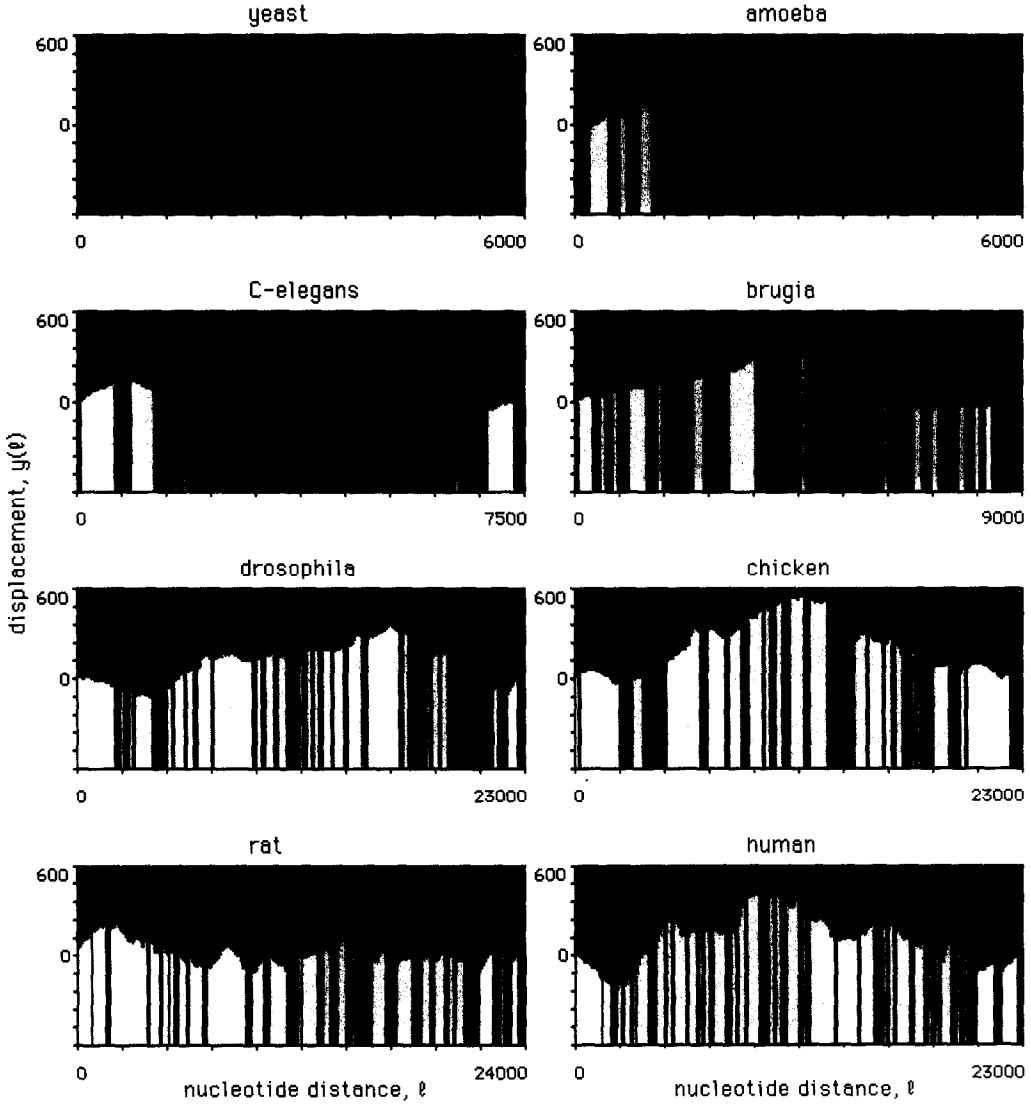
yeast
amoeba
C-elegans
brugia
drosophila
chicken
rat
human
600
0
0
6000
7500
9000
23000
24000
displacement, y(ℓ)
nucleotide distance, ℓ

Fig. 12b.

puzzle, then, is how hydrochloric acid, secreted at the base of gastric glands by specialized parietal cells, traverses the mucus layer to reach the lumen without acidifying the mucus layer. Bhaskar et al. [15] resolved this puzzle by experiments that demonstrate the the possibility that flow of hydrochloric acid through mucus involves viscous fingering—the phenomenon that occurs when a fluid of lower viscosity is injected into a more viscous one (see Fig. 3). Specifically, Bhaskar et al. demonstrated that injection of hydrochloric acid through solutions of pig gastric mucin produces fingering patterns which are strongly dependent on pH, mucin concentration, and acid flow rate. Above pH 4, discrete fingers are observed, while below pH 4, hydrochloric acid neither penetrates the mucin solution nor forms fingers. These *in vitro* results suggest that hydrochloric acid secreted by the gastric gland can penetrate the mucus gel layer (pH 5-7) through narrow fingers, whereas hydrochloric acid in the lumen (pH 2) is prevented from diffusing back to the epithelium by the high viscosity of gastric mucus gel on the luminal side.

Yet another example of DLA-type growth is bacterial colony spread on the surface of agar (gel with nutrient) plates [16] (see Fig. 4). Vicsek et al. [17] studied bacterial colony growth on a strip geometry which results in a self-affine surface (see Fig. 13.19 in [18]). They calculated the roughness exponent α for this surface and found $\alpha = 0.78 \pm 0.07$. The interfacial pattern formation of the growth of bacterial colonies was studied systematically by Ben-Jacob et al. [19]. They demonstrated that bacterial colonies can develop a pattern similar to morphologies in diffusion-limited growth observed in solidification and electro-chemical deposition. These include fractal growth, dense-branching growth, compact growth, dendritic growth and chiral growth. The results indicate that the interplay between the micro level (individual bacterium) and the macro level (the colony) play a major role in selecting the observed morphologies similar to those found in nonliving systems.

Another example of fractal interface appears in ecology, in the problem of the territory covered by N diffusing particles [20], see Fig. 5. As seen from the figure, the territory initially grows with the shape of a disk with a relatively smooth surface until it reaches a certain size, at which point the surface becomes increasingly rough. This phenomenon may have been observed by Skellam [21] who plotted contours delineating the advance of the muskrat population and noted that initially the contours were smooth but at later times they became rough (see Fig. 1 in [21]).

Other biological contexts in which fractal scaling seems to be relevant are the relation between brain size and body weight [22], between bone diameter and bone length [23], between muscle force and muscle mass [23], and between an organism's size and its rate of producing energy and consuming food [24].

3 Long-Range Power Law Correlations

In recent years long-range power-law correlations have been discovered in a remarkably 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" [1,3,18,25–28]. So if fractals are indeed so widespread, it makes sense to anticipate that long-range power-law correlations may be similarly widespread. Indeed, 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 for apparently unrelated systems allows us to recognize similarities between different systems, leading to underlying unifications that might otherwise have gone unnoticed.

Usually correlations decay exponentially, but there is one major exception: at the critical point [29], the exponential decay of (1a) turns into to a power law decay

$$C_r \sim (1/r)^{d-2+\eta}. \tag{1}$$

Many systems drive themselves spontaneously toward critical points [30,31]. One of the simplest models exhibit-

ing such "self-organized criticality" is invasion percolation, a generic model that has recently found applicability to describing anomalous behavior of rough interfaces. Instead of occupying all sites with random numbers below a pre-set parameter p, in invasion percolation one "grows" the incipient infinite cluster right at the percolation threshold by the trick of occupying always the perimeter site whose random number is smallest. Thus small clusters are certainly not scale-invariant and in fact contain sites with a wide distribution of random numbers. As the mass of the cluster increases, the cluster becomes closer and closer to being scale invariant or "fractal." Such a system is said to drive itself to a "self-organized critical state" [32].

In the following sections we will attempt to summarize the key findings of some recent work [33–57] suggesting that—under suitable conditions—the sequence of base pairs or "nucleotides" in DNA also displays power-law correlations. The underlying basis of such power law correlations is not understood at present, but this discovery has intriguing implications for molecular evolution and DNA structure, as well as potential practical applications for distinguishing coding and noncoding regions in long nucleotide chains.

4 Information Coding in DNA

The role of genomic DNA sequences in coding for protein structure is well known [58,59]. The human genome contains information for approximately 100,000 different proteins, which define all inheritable features of an individual. The genomic sequence is likely the most sophisticated information database created by nature through the dynamic process of evolution. Equally remarkable is the precise transformation of information (duplication, decoding, etc) that occurs in a relatively short time interval.

The building blocks for coding this information are called *nucleotides*. Each nucleotide contains a phosphate group, a deoxyribose sugar moiety and either a *purine* or a *pyrimidine base*. Two purines and two pyrimidines are found in DNA. The two purines are adenine (A) and guanine (G); the two pyrimidines are cytosine (C) and thymine (T). The nucleotides are linked end to end, by chemical bonds from the phosphate group of one nucleotide to the deoxyribose sugar group of the adjacent nucleotide, forming a long polymer (*polynucleotide*) chain. The information content is encoded in the sequential order of the bases on this chain. Therefore, as far as the information content is concerned, a DNA sequence can be most simply represented as a symbolic sequence of four letters: A, C, G and T, as shown in Fig. 6.

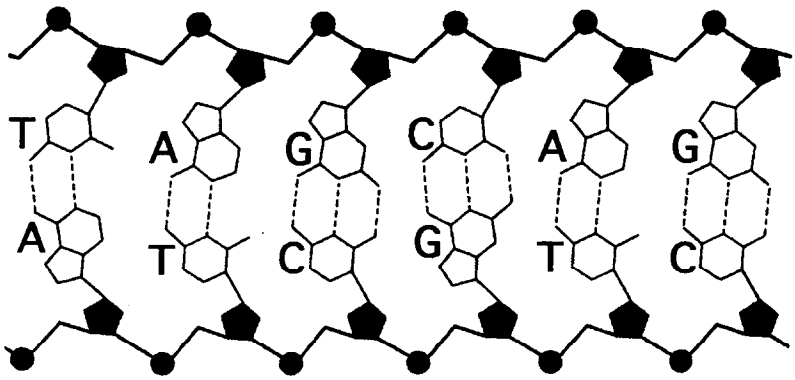

Fig. 6. The base pairing of two "double helix" DNA strands. The two chains of black pentagons and circles represent sugar-phosphate backbones of DNA strands linked by the hydrogen bonds (dashed lines) between complementary base pairs

In the genomes of high eukaryotic organisms only a small portion of the total genome length is used for protein coding (as low as 5% in the human genome). For example, genes are separated from each other by *intergenic sequences* which are not used for coding proteins and which (especially in mammalian genomes) can be several times longer than genes. Furthermore, in 1977 it was discovered that genes themselves have inclusions which are not used for coding proteins. A gene is transcripted to RNA (pre-mRNA) and then some segments of the pre-mRNA are "spliced out" during the formation of the smaller mRNA molecule. The mRNA then serves as the template for assembling protein. 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 intergenomic 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.

5 Conventional Statistical Analysis of DNA Sequences

DNA sequences have been analyzed using a variety of models that can basically be considered in two categories. The first types are "local" analyses; they take into account the fact that DNA sequences are produced in sequential order; therefore, the neighboring nucleotides will affect the next attaching nucleotide. This type of analysis, represented by n-step Markov models, can indeed describe some observed short-range correlations in DNA sequences. The second category of analyses is more "global" in nature; they concentrate on the presence of repeated patterns (such as periodic repeats and interspersed base sequence repeats) that are chiefly found in eukaryotic genomic sequences. A typical example of analysis in this category is the Fourier transform, which can identify repeats of certain segments of the same length in nucleotide sequences [58].

However, DNA sequences are more complicated than these two standard types of analysis can describe. Therefore it is crucial to develop new tools for analysis with a view toward uncovering the mechanisms used to code other types of information. Promising techniques for genome studies may be derived from other fields of scientific research, including time-series analysis, statistical mechanics, fractal geometry, and even linguistics.

6 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

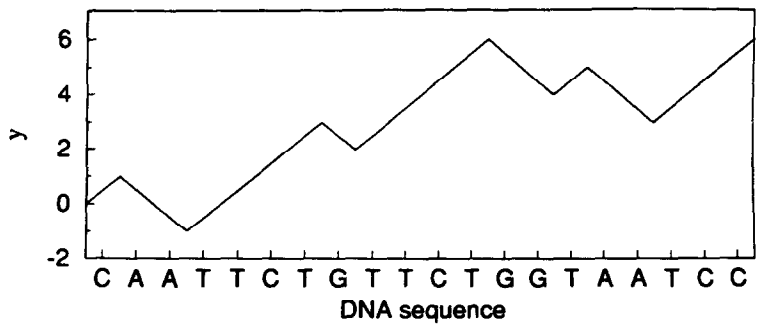

Fig. 7. Schematic illustration showing the definition of the "DNA walk"

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 [29].

6.1 Graphical Representation

In order to study the scale-invariant long-range correlations of a DNA sequence, we first introduced a graphical representation of DNA sequences, which we term a *fractal landscape* or *DNA walk* [33]. For the conventional one-dimensional random walk model [60,61], 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 [62–64].

One definition of the DNA walk is that the walker steps "up" $[u(i) = +1]$ if a pyrimidine (C or T) occurs at position i along the DNA chain, while the walker steps "down" $[u(i) = -1]$ if a purine (A or G) occurs at position i (see Fig. 7). 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).

There have also been attempts to map DNA sequence onto multi-dimensional DNA walks [34,65]. However, recent work [57] indicates that the original purine-pyrimidine rule provides the most robust results, probably due to the purine-pyrimidine chemical complementarity.

The DNA walk allows one to visualize directly the fluctuations of the purine-pyrimidine content in DNA sequences: Positive slopes on Fig. 8 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. Figure 8a shows a typical example of a gene that contains a significant fraction of base pairs that do *not* code for amino acids. Figure 8b shows the DNA walk for a sequence formed by splicing together the coding regions of the DNA sequence of this same gene (i.e., the cDNA). Figure 8c displays the DNA walk for a typical sequence with only coding regions. Landscapes for intron-containing sequences show very jagged contours which consist of patches of all length scales, reminiscent of the disordered state of matter near critical point. On the other hand, coding sequences typically consist of a few lengthy regions of different strand bias, resembling domains in the system in the ferromagnet state. These observations can be tested by rigorous statistical analysis. Figure 8 naturally motivates a quantification of these fluctuations by calculating the "net displacement" of the walker after ℓ steps, which is the sum of the unit steps $u(i)$ for each step i. Thus $y(\ell) \equiv \sum_{i=1}^{\ell} u(i)$.

6.2 Correlations and Fluctuations

An important statistical quantity characterizing any walk [60,61] is the root mean square fluctuation $F(\ell)$ about the average of the displacement; $F(\ell)$ is defined in terms of the difference between the average of the square and the square of the average,

$$F^2(\ell) \equiv \overline{[\Delta y(\ell) - \overline{\Delta y(\ell)}]^2} = \overline{[\Delta y(\ell)]^2} - \overline{\Delta y(\ell)}^2, \tag{2}$$

of a quantity $\Delta y(\ell)$ defined by $\Delta y(\ell) \equiv y(\ell_0 + \ell) - y(\ell_0)$ (see also Chaps. 1 and 5). Here the bars indicate an

average over all positions ℓ_0 in the gene. Operationally, this is equivalent to (a) using calipers preset for a fixed distance ℓ, (b) moving the beginning point sequentially from $\ell_o = 1$ to $\ell_o = 2, \cdots$ and (c) calculating the quantity $\Delta y(\ell)$ (and its square) for each value of ℓ_o, and (d) averaging all of the calculated quantities to obtain $F^2(\ell)$.

The mean square fluctuation is related to the auto-correlation function

$$C(\ell) \equiv \overline{u(\ell_0)u(\ell_0+\ell)} - \overline{u(\ell_0)}^2 \quad \text{through the relation} \quad F^2(\ell) = \sum_{i=1}^{\ell}\sum_{j=1}^{\ell} C(j-i). \tag{3}$$

The calculation of $F(\ell)$ can distinguish three possible types of behavior.

(i) If the base pair sequence were random, then $C(\ell)$ would be zero on average [except $C(0) = 1$], so $F(\ell) \sim \ell^{1/2}$ (as expected for a *normal* random walk).

(ii) If there were local correlations extending up to a characteristic range R (such as in Markov chains), then $C(\ell) \sim \exp(-\ell/R)$; nonetheless *the asymptotic* ($\ell \gg R$) *behavior* $F(\ell) \sim \ell^{1/2}$ *would be unchanged from the purely random case.*

(iii) If there is no characteristic length (i.e., if the correlation were "infinite-range"), then the scaling property of $C(\ell)$ would not be exponential, but would most likely to be a power law function, and the fluctuations will also be described by a power law

$$F(\ell) \sim \ell^{\alpha}, \tag{4}$$

with $\alpha \neq 1/2$.

Figure 8a shows a typical example of a gene that contains a significant fraction of base pairs that do *not* code for amino acids. It is immediately apparent that the DNA walk has an extremely jagged contour which corresponds to long-range correlations. Figure 9 shows double logarithmic plots of the mean square fluctuation function $F(\ell)$ as a function of the linear distance ℓ along the DNA chain for a typical intron-containing gene.

The fact that the 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 the prediction for an uncorrelated walk, $\alpha = 1/2$, thus providing direct experimental evidence for the presence of long-range correlations.

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.

It is known that functional proteins usually form a single compact three-dimensional conformation that corresponds to the global energy minimum in the conformational space. Recently, Shakhnovich and Gutin [66] found that in order to have such a minimum it is sufficient that an amino acid sequence forms an uncorrelated random sequence. The finding of Peng et al. [33] of the lack of long range correlations in the coding nucleotide sequences provides more evidence for this hypothesis, since there exist almost one-to-one correspondence between amino acid sequences and their nucleotide codes. Furthermore, this finding may also indicate that the *lack* of long range correlations in the amino acid sequences is, in fact, a necessary condition for a functional biologically active protein.

7 Differences Between Correlation Properties of Coding and Noncoding Regions

The initial report [33] on long-range (scale-invariant) correlations only in noncoding DNA sequences has gener-

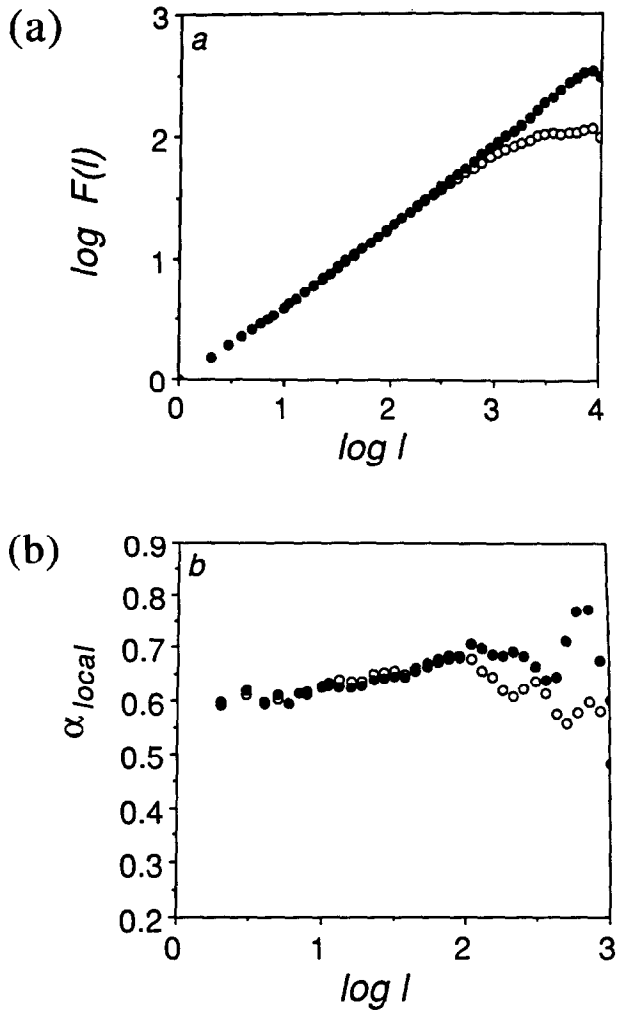

Fig. 9. (a) Double logarithmic plots of the mean square fluctuation function $F(\ell)$ as a function of the linear distance ℓ along the DNA chain for the rat embryonic skeletal myosin heavy chain gene (○) and its "intron-spliced sequence" (•). (b) The corresponding local slopes, α_{local}, based on pairs of successive data points of part (a). We see that the values of α are roughly constant. For this specific gene, the sequence with exons removed has an even broader scaling regime than the DNA sequence of the entire gene, indicated by the fact that part (a) is linear up to 10,000 nucleotides. After [42]

ated contradicting responses. Some [34,35, 38,39] support our initial finding, while some [35,40,44,50] disagree. However, the conclusions of Refs. [36] and [35,40,44,50] are inconsistent *with one another* in that [35] and [50] doubt the existence of long-range correlations (even in noncoding sequences) while [36] and [40,44] conclude that even coding regions display long-range correlations ($\alpha > 1/2$). Prabhu and Claverie [40] claim that their analysis of the putative *coding* regions of the yeast chromosome III [67] 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. Hence conventional scaling analyses cannot be applied reliably to the entire sequence but only to sub-sequences.

Peng et al. [56] have recently applied the "bridge method" to DNA, and have also developed a similar method

specifically adapted to handle problems associated with non-stationary 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. One source of this difference is the tandem repeats (sequences such as AAAAAA...), which are quite frequent in noncoding sequences and absent in the coding sequences.

To provide an "unbiased" test of the thesis that noncoding regions possess but coding regions lack long-range correlations, Ossadnik et al. [57] analyzed several artificial uncorrelated and correlated "control sequences" of size 10^5 nucleotides using the GRAIL neural net algorithm [68]. The GRAIL algorithm identified about 60 putative exons in the uncorrelated sequences, but only about 5 putative exons in the correlated sequences.

Using the DFA method, we can measure the local value of the correlation exponent α along the sequence (see Fig 10) and find that the local minima of α as a function of a nucleotide position usually correspond to noncoding regions, while the local maxima correspond to noncoding regions. Statistical analysis using the DFA technique of the nucleotide sequence data for yeast chromosome III (315,338 nucleotides) shows that the probability that the observed correspondence between the positions of minima and coding regions is due to random coincidence is less than 0.0014. Thus, this method—which we called the "coding sequence finder" (CSF) algorithm—can be used for finding coding regions in the newly sequenced DNA, a potentially important application of DNA walk analysis.

8 Long-Range Correlations and Evolution

What is the biological meaning of the finding of long-range correlations in DNA? If two nucleotides whose positions differ by 1000 base pairs were uncorrelated, then there might be no meaning. However, the finding that they are correlated suggests some underlying organizational property. The long-range correlations in DNA sequences are of interest because they may be an indirect clue to its three-dimensional structure [45,54] or a reflection of certain scale-invariant properties of long polymer chains [53,55]. In any case, the statistically meaningful long-range "scale-invariant" (see Fig. 11) correlations in noncoding regions and their absence in coding regions will need to be accounted for by future explanations of global properties in gene organization and evolution.

Molecular evolutionary relationships are usually inferred from comparison of coding sequences, conservation of intron/exon structure of related sequences, analysis of nucleotide substitutions, and construction of phylogenetic trees [69]. The changes observed are conventionally interpreted with respect to nucleotide sequence composition (mutations, deletions, substitutions, alternative splicing, transpositions, etc.) rather than overall genomic organization.

Very recently, Buldyrev et al. [55] sought to assess the utility of DNA correlation analysis as a complementary method of studying gene evolution. In particular, they studied the changes in "fractal complexity" of nucleotide organization of a single gene family with evolution. A recent study by Voss [36] reported that the correlation ex-

ponent derived from Fourier analysis was lowest for sequences from organelles, but paradoxically higher for invertebrates than vertebrates. However, this analysis must be interpreted with caution since it was based on pooled data from different gene families rather than from the quantitative examination of any single gene family (see also [70,71]).

The hypothesis that the fractal complexity of genes from higher animals is greater than that of lower animals, using single gene family analysis was tested in [55]. This analysis focuses on the genome sequences from the conventional (Type II) myosin heavy chain (MHC) family. Such a choice limits potential bias that may arise secondary to non-uniform evolutionary pressures and differences in nucleotide content between unrelated genes. The MHC gene family was chosen because of the availability of completely sequenced genes from a phylogenetically diverse group of organisms, and the fact that their relatively long sequences are well-suited to statistical analysis.

The landscape produced by DNA walk analysis reveals that each MHC cDNA consists of two roughly equal parts with significant differences in nucleotide content (Fig. 12). The first part that codes for the heavy meromyosin or "head" of the protein molecule has a slight excess of purines (52% purines and 48% pyrimidines); the second part that codes for the light meromyosin or "tail" has about 63% purines and 37% pyrimidines. The *absolute nucleotide* contents are not shown in the graphical representation of Fig. 12a because we subtract the average slope from the landscape to make relative fluctuations around the average more visible. Indeed, one can easily see from Fig. 12a that the relative concentration of pyrimidines in the first part ("uphill" region) of the myosin cDNA is much higher than in the second ("downhill" region).

The landscapes of Fig. 12 show that the coding sequences of myosins remain practically unchanged with evolution, while the entire gene sequences become more heterogeneous and complex. The quantitative measurements of the exponent α by DFA method confirm this visual observation showing that for all coding sequences of MHC family $\alpha \approx 0.5$. In contrast, for entire genes of MHC family, the value of α monotonically increases from lower eukaryotes to invertebrates and from invertebrates to vertebrates [55]. A stochastic model of random deletions and insertions of DNA portions was developed in [55] to explain this finding; see also [72–77].

Two major theories have been advanced to explain the origin and evolution of introns. One suggests that precursor genes consisted entirely of coding sequences and introns were inserted later in the course of evolution to help facilitate development of new structures in response to selective pressure, perhaps, by means of "exon shuffling" [78]. The alternative theory suggests that precursor genes were highly segmented and subsequently organisms not requiring extensive adaptation or new development or, perhaps, facing the high energetic costs of replicating unnecessary sequences, lost their introns [79,80]. Support for these hypotheses has remained largely conjectural; no models have been brought forward to support either process. The landscape analysis of the MHC gene family and the stochastic model [53,55] here are more consistent with the former view.

9 Other Biological Systems with Long-Range Correlations

The catalog of systems in which power law correlations appear has grown rapidly in recent years [32,81,82]; see also Chap. 2. What do we anticipate for other biological systems? Generally speaking, when "entropy wins over energy"—i.e., randomness dominates the behavior—we find power laws and scale invariance. The absence of characteristic length (or time) scales may confer important biological advantages, related to adaptability of response [2]. Biological systems sometimes are described in language that makes one think of a Swiss watch. Such mechanistic or "Rube Goldberg" descriptions must in some sense be incomplete, since it is only some appropriately-chosen averages that appear to behave in a regular fashion. The trajectory of each individual biological molecule is of necessity random—albeit correlated. Thus one might hope that recent advances in understanding "correlated randomness" [83,62–64] could be relevant to biological phenomena.

9.1 *The Human Heartbeat*

Traditionally, clinicians describe the normal electrical activity of the heart as "regular sinus rhythm." However, cardiac interbeat intervals fluctuate in a complex, apparently erratic manner in healthy subjects even at rest. Analysis of heart rate variability has focused primarily on short time oscillations associated with breathing (0.15–0.40 Hz) and blood pressure control ($\sim$ 0.1 Hz) [80]. Fourier analysis of longer heart rate data sets from healthy individuals typically reveals a $1/f$-like spectrum for frequencies < 0.1 Hz [84–87].

Peng et al. [88] recently studied scale-invariant properties of the human heartbeat time series, the output of a complicated integrative control system. The analysis is based on the digitized electrocardiograms of beat-to-beat heart rate fluctuations over very long time intervals (up to 24 h $\approx 10^5$ beats) recorded with an ambulatory monitor. The time series obtained by plotting the sequential intervals between beat n and beat $n+1$, denoted by $B(n)$, typically reveals a complex type of variability. The mechanism underlying such fluctuations is related to competing neuroautonomic inputs. Parasympathetic (vagal) stimulation decreases the firing rate of pacemaker cells in the heart's sinus node; sympathetic stimulation has the opposite effect. The nonlinear interaction (competition) between these two branches of the involuntary nervous system is the postulated mechanism for much of the erratic heart rate variability recorded in healthy subjects, although non-autonomic factors may also be important.

To study these dynamics over large time scales, the time series is passed through a digital filter that removes fluctuations of frequencies > 0.005 beat^{-1}, and plot the result, denoted by $B_L(n)$, in Fig. 13. One observes a more complex pattern of fluctuations for a representative healthy adult (Fig. 13a) compared to the "smoother" pattern of interbeat intervals for a subject with severe heart disease (Fig. 13b). These heartbeat time series produce a contour reminiscent of the irregular landscapes that have been widely studied in physical systems.

To quantitatively characterize such a "landscape," Peng et al. introduce a mean fluctuation function $F(n)$, defined as

$$F(n) \equiv \overline{|B_L(n'+n) - B_L(n')|}, \tag{5}$$

where the bar denotes an average over all values of n'. Since $F(n)$ measures the average difference between two interbeat intervals separated by a time lag n, $F(n)$ quantifies the magnitude of the fluctuation over different time scales n.

Figure 13c is a log-log plot of $F(n)$ vs n for the data in Figs. 13a and 13b. This plot is approximately linear over a broad physiologically-relevant time scale (200 – 4000 beats) implying that

$$F(n) \sim n^{\alpha}. \tag{6}$$

It is found that the scaling exponent α is markedly different for the healthy and diseased states: for the healthy heartbeat data, α is close to 0, while α is close to 0.5 for the diseased case. Note that $\alpha = 0.5$ corresponds to a random walk (a Brownian motion), thus the low-frequency heartbeat fluctuations for a diseased state can be interpreted as a stochastic process, in which the heartbeat intervals $I(n) \equiv B(n+1) - B(n)$ are uncorrelated for $n \geq 200$.

To investigate these dynamical differences, it is helpful to study further the correlation properties of the time series. It is useful to study $I(n)$ because it is the appropriate variable for the aforementioned reason. Since $I(n)$ is stationary, one can apply standard spectral analysis techniques [3] (see also Chap. 1). Figures 14a and 14b show the power spectra $S_I(f)$, the square of the Fourier transform amplitudes for $I(n)$, derived from the same data sets (without filtering) used in Fig. 13. The fact that the log-log plot of $S_I(f)$ vs f is linear implies

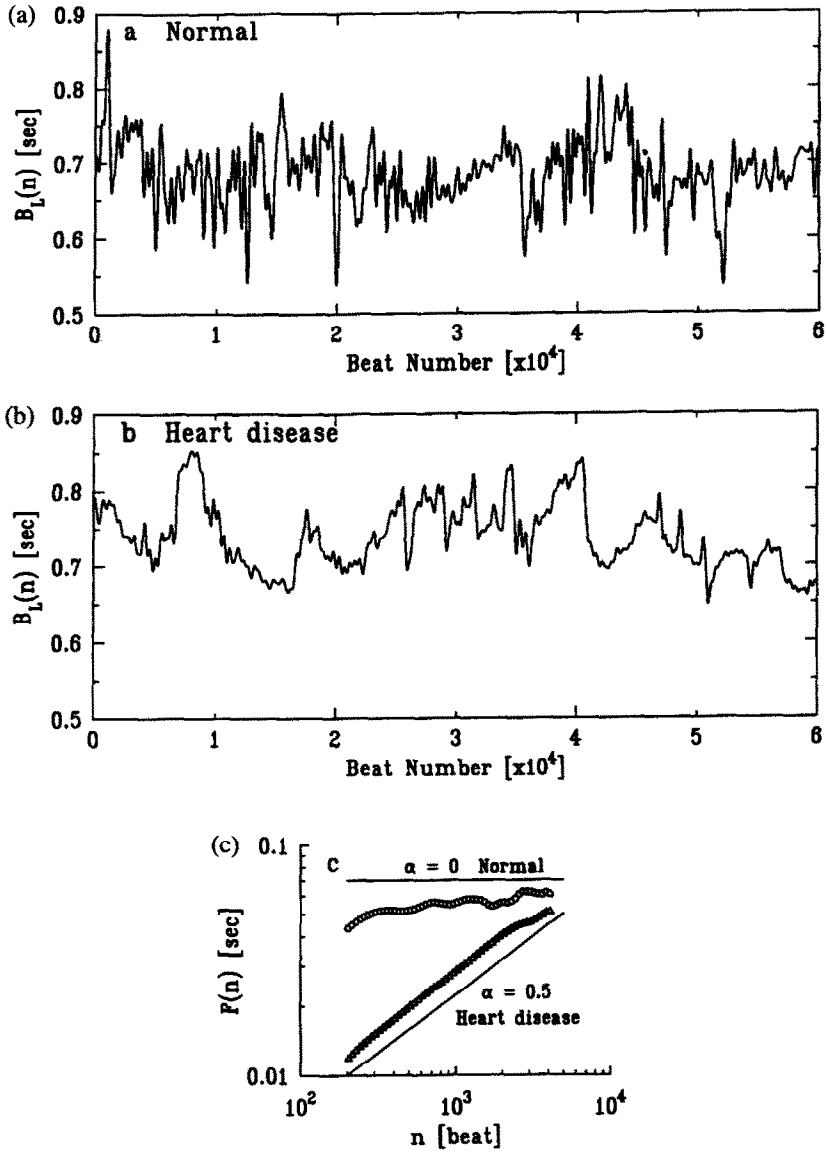

Fig. 13. The interbeat interval $B_L(n)$ after low-pass filtering for (a) a healthy subject and (b) a patient with severe cardiac disease (dilated cardiomyopathy). The healthy heartbeat time series shows more complex fluctuations compared to the diseased heart rate fluctuation pattern that is close to random walk ("brown") noise. (c) Log-log plot of $F(n)$ vs n. The circles represent $F(n)$ calculated from data in (a) and the triangles from data in (b). The two best-fit lines have slope $\alpha = 0.07$ and $\alpha = 0.49$ (fit from 200 to 4000 beats). The two lines with slopes $\alpha = 0$ and $\alpha = 0.5$ correspond to "$1/f$ noise" and "brown noise," respectively. We observe that $F(n)$ saturates for large n (of the order of 5000 beats), because the heartbeat interval are subjected to physiological constraints that cannot be arbitrarily large or small. The low-pass filter removes all Fourier components for $f \geq f_c$. The results shown here correspond to $f_c = 0.005$ beat^{-1}, but similar findings are obtained for other choices of $f_c \leq 0.005$. This cut-off frequency f_c is selected to remove components of heart rate variability associated with physiologic respiration or pathologic Cheyne-Stokes breathing as well as oscillations associated with baroreflex activation (Mayer waves). After [88]

$$S_I(f) \sim \frac{1}{f^{\beta}}. \tag{7}$$

The exponent β is related to α by $\beta = 2\alpha - 1$ [62]. Furthermore, β can serve as an indicator of the presence and type of correlations:

(i) If $\beta = 0$, there is no correlation in the time series $I(n)$ ("white noise").

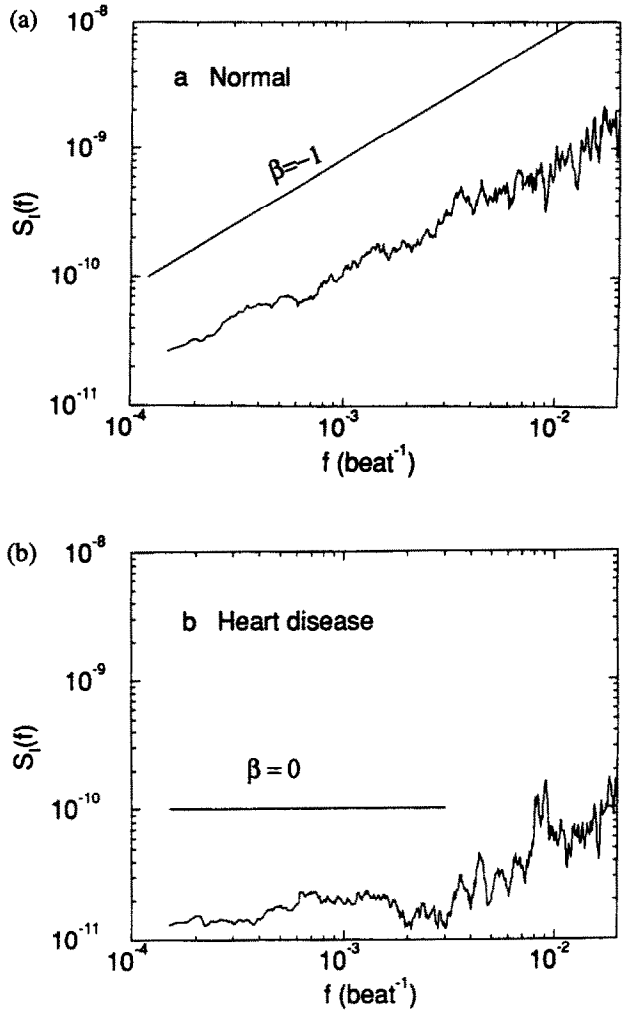

Fig. 14. The power spectrum $S_I(f)$ for the interbeat interval increment sequences over $\sim$ 24 hours for the same subjects in Fig. 13. (a) Data from a healthy adult. The best-fit line for the low frequency region has a slope $\beta = -0.93$. The heart rate spectrum is plotted as a function of "inverse beat number" (beat^{-1}) rather than frequency (time^{-1}) to obviate the need to interpolate data points. The spectral data are smoothed by averaging over 50 values. (b) Data from a patient with severe heart failure. The best-fit line has slope 0.14 for the low frequency region, $f < f_c = 0.005$ beat^{-1}. The appearance of a pathologic, characteristic time scale is associated with a spectral peak (arrow) at about 10^{-2} beat^{-1} (corresponding to Cheyne-Stokes respiration). After [88]

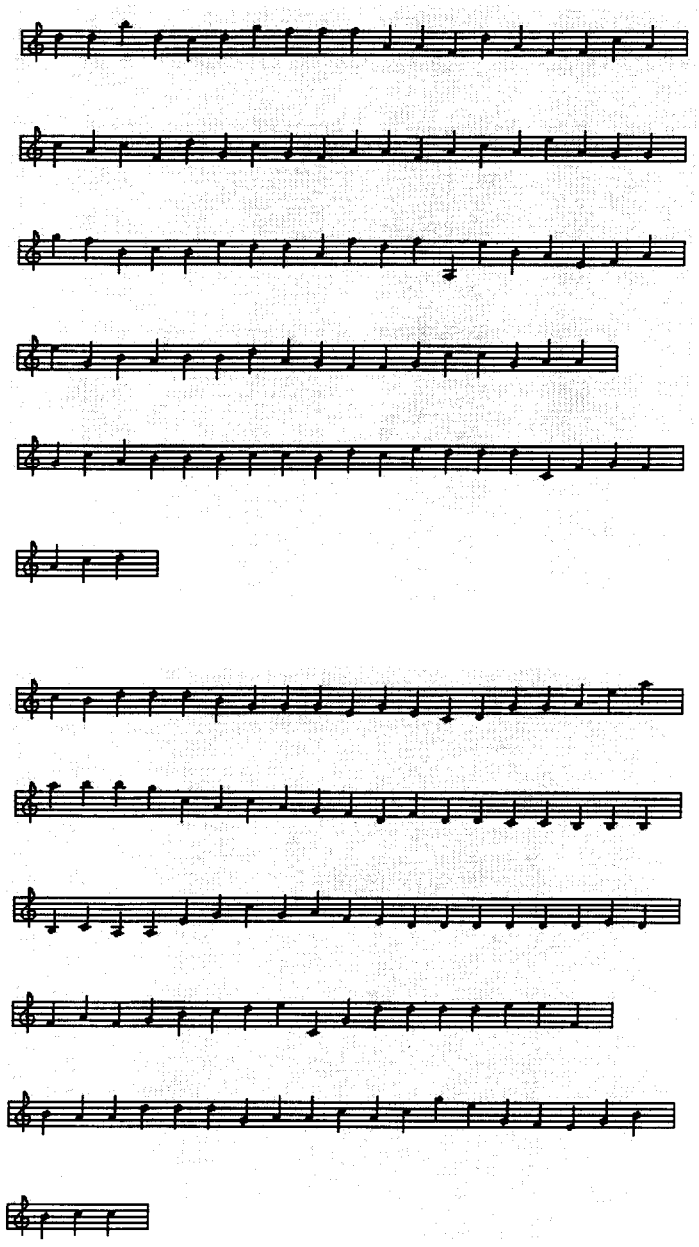

Fig. 15. Musical mapping of two heartbeat times series, derived from normal (top) and pathologic (bottom) data sets. The original heart beat time series were obtained from 24 hour recordings consisting of about 10^6 heartbeats. The heartbeat time series were then low-pass filtered to remove fluctuations > 0.05 (beat^{-1}), roughly equivalent to averaging every 200 beats. The pattern of fluctuations in the normal is more complex than that of the "music" generated from the abnormal data sets. Musical compositions based on these times series are available on cassette by request along with the "scores"; contact Zachary D. Goldberger (e-mail: ary "at" astro.bih.harvard.edu). There is a nominal charge for copying and mailing. After [97], courtesy of Z.D. Goldberger

(ii) If $0 < \beta < 1$, then $I(n)$ is correlated such that positive values of I are likely to be close (in time) to each other, and the same is true for negative I values.
(iii) If $-1 < \beta < 0$, then $I(n)$ is also correlated; however, the values of I are organized such that positive and negative values are more likely to alternate in time ("anti-correlation") [62].

For the diseased data set, we observe a flat spectrum ($\beta \approx 0$) in the low frequency region (Fig. 14b) confirming that $I(n)$ are not correlated over long time scales (low frequencies). Therefore, $I(n)$, the first derivative of $B(n)$, can be interpreted as being analogous to the *velocity* of a random walker, which is uncorrelated on long time scales, while $B(n)$—corresponding to the *position* of the random walker—are correlated. However, this correlation is of a trivial nature since it is simply due to the summation of uncorrelated random variables.

In contrast, for the data set from the healthy subject (Fig. 14a), we obtain $\beta \approx -1$, indicating *nontrivial* long-range correlations in $B(n)$—these correlations are not the consequence of summation over random variables or artifacts of non-stationarity. Furthermore, the "anti-correlation" properties of $I(n)$ indicated by the negative β value are consistent with a nonlinear feedback system that "kicks" the heart rate away from extremes. This tendency, however, does not only operate on a beat-to-beat basis (local effect) but on a wide range of time scales. To our knowledge, this is the first explicit description of long-range anticorrelations in a fundamental biological variable, namely the interbeat interval increments.

Furthermore, the histogram for the heartbeat intervals increments is found to be well-described by a Lévy stable distribution. For a group of subjects with severe heart disease, it is found that the distribution is unchanged, but the long-range correlations vanish. Therefore, the different scaling behavior in health and disease must relate to the underlying dynamics of the heartbeat. Applications of this analysis may lead to new diagnostics for patients at high risk of cardiac disease and sudden death.

9.2 *Physiological Implications*

The finding of nontrivial long-range correlations in healthy heart rate dynamics is consistent with the observation of long-range correlations in other biological systems that do not have a characteristic scale of time or length. Such behavior may be adaptive for at least two reasons. (i) The long-range correlations serve as an organizing principle for highly complex, nonlinear processes that generate fluctuations on a wide range of time scales. (ii) The lack of a characteristic scale helps prevent excessive *mode-locking* that would restrict the functional responsiveness of the organism. Support for these related conjectures is provided by observations from severe diseased states such as heart failure where the breakdown of long-range correlations is often accompanied by the emergence of a dominant frequency mode (e.g., the Cheyne-Stokes frequency). Analogous transitions to highly periodic regimes have been observed in a wide range of other disease states including certain malignancies, sudden cardiac death, epilepsy and fetal distress syndromes.

The complete breakdown of normal long-range correlations in any physiological system could theoretically lead to three possible diseased states: (i) a random walk (brown noise), (ii) highly periodic behavior, or (iii) completely uncorrelated behavior (white noise). Cases (i) and (ii) both indicate only "trivial" long-range correlations of the types observed in severe heart failure. Case (iii) may correspond to certain cardiac arrhythmias such as fibrillation. More subtle or intermittent degradation of long-range correlation properties may provide an early warning of incipient pathology. Finally, we note that the long-range correlations present in the healthy heartbeat indicate that the neuroautonomic control mechanism actually drives the system away from a single steady state. Therefore, the classical theory of homeostasis, according to which stable physiological processes seek to maintain "constancy" [89], should be extended to account for this dynamical, far from equilibrium, behavior.

9.3 Human Writings

Long-range correlations have been found recently in human writings [81]. A novel, a piece of music or a computer program can be regarded as a one-dimensional string of symbols. These strings can be mapped to a one-dimensional random walk model similar to the DNA walk (Sect. 6) allowing calculation of the correlation exponent α using (4a). Values of α between 0.6 and 0.9 were found for various texts.

An interesting fractal feature of languages was found in 1949 by Zipf [90]. He observed that the frequency of words as a function of the word order decays as a power law (with a power close to -1) for more than four orders of magnitude. A theory for this empirical finding, based on assumptions of coding words in the brain, was given by Mandelbrot [1,91]. A related interesting statistical measure of short-range correlations in languages and in general series sequences is the entropy and redundancy defined by Shannon [92].

9.4 Dynamics of Membrane Channel Openings

Ions, such as potassium and sodium, cannot cross the lipid cell membrane. They can, however, enter or exit the cell through ion channel proteins that are distributed on the cell membrane. These proteins spontaneously fluctuate between open or closed states. Liebovitch [93] found that the histograms of the open and closed duration times of some channels are self-similar and behave as power laws. This approach may provide new models for the ion channel gating mechanisms.

9.5 Fractal Music and the Heartbeat

Fourier analysis of instantaneous fluctuations in amplitude as well as inter-note intervals for certain classical music pieces (e.g., Bach's First Brandenburg Concerto) reveals a $1/f$ distribution over a broad frequency range [94-96]. Voss and Clarke [95] used an algorithm for generating $1/f$-noise to "compose" music.

Based on the observation of different scaling patterns for healthy and pathologic heartbeat time series [88], it was very recently postulated [97] that (i) actual biological rhythms such as the heartbeat might serve as a more natural template for musical compositions than artificially-generated noises, and (ii) audibly appreciable differences between the note series of healthy and diseased hearts could potentially serve as the basis for a clinically useful diagnostic test. Accordingly, a computer program was devised to map heart rate fluctuations onto intervals of the diatonic musical scale [97]. As anticipated, the *normal* ($1/f$-like) heartbeat obtained from the low pass filtered time series reported in [88] generated a more variable (complex) type of music than that generated by the *abnormal* times series (Fig. 15).

The "musicality" of these transcriptions is intriguing and supports speculations about the brain's possible role as a translator/manipulator of biological $1/f$-like noise into aesthetically pleasing art works. Current investigations are aimed at extending these preliminary observations by (i) comparing the "musicality" of note sequences generated by natural (biological) vs. artificial (computer simulated) correlated and uncorrelated noises, and (ii) using heartbeat time series as the template for simultaneously generating fluctuations in musical rhythm and intensity, not only pitch.

9.6 Fractal Approach to Biological Evolution

Fossil data suggest that evolution of biological species takes place as intermittent bursts of activity, separated

by relatively long periods of quiescence [98]. Recently Bak and Sneppen [99] suggested that these spontaneous catastrophic extinctions may be related to the power law distribution of avalanches of growth observed in a model of self-organized criticality (SOC). Such SOC models are reminiscent of recent surface growth models based on the concept of directed invasion percolation [100].

Acknowledgements

We are grateful to R. Bansil, A.-L. Barabasi, K.R. Bhaskar, F. Caserta, G. Daccord, W. Eldred, P. Garik, Z.D. Goldberger, Z. Hantos, J.M. Hausdorff, R.E. Hausman, P. Ivanov, T.J. LaMont, H. Larralde, M.E. Matsa, J. Mietus, A. Neer, J. Nittmann, F. Pétak, I. Rabin, F. Sciortino, A. Shehter, M.H.R. Stanley, B. Suki, P. Trunfio, M. Uideja, and G.H. Weiss 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, G. Huber, I. Labat, L. Liebovitch, G.S. Michaels, P. Munson, R. Nossal, R. Nussinov, R.D. Rosenberg, J.J. Schwartz, M. Schwartz, E.I. Shakhnovich, M.F. Shlesinger, N. Shworak, and E.N. Trifonov for valuable discussions. Partial support was provided to to SVB and HES by the National Science Foundation, to ALG by the G. Harold and Leila Y. Mathers Charitable Foundation, the National Heart, Lung and Blood Institute and the National Aeronautics and Space Administration, to SH by Israel-USA Binational Science foundation, and to C-KP by an NIH/NIMH Postdoctoral NRSA Fellowship.

References

[1] B.B. Mandelbrot: *The Fractal Geometry of Nature* (W.H. Freeman, San Francisco 1982)

[2] B.J. West, A.L. Goldberger: J. Appl. Physiol., **60**, 189 (1986); B.J. West, A.L. Goldberger: Am. Sci., **75**, 354 (1987); A.L. Goldberger, B.J. West: Yale J. Biol. Med. **60**, 421 (1987); A.L. Goldberger, D.R. Rigney, B.J. West: Sci. Am. **262**, 42 (1990); B.J. West, M.F. Shlesinger: Am. Sci. **78**, 40 (1990); B.J. West: *Fractal Physiology and Chaos in Medicine* (World Scientific, Singapore 1990); B.J. West, W. Deering: Phys. Reports **xx**, xxx (1994)

[3] A. Bunde, S. Havlin, eds.: *Fractals and Disordered Systems* (Springer-Verlag, Berlin 1991); see also J. Kertesz and T. Vicsek, in A. Bunde and S. Havlin, eds.: *Fractals in Science* (Springer-Verlag, Berlin 1994) and A.-L. Barabasi and H. E. Stanley, *Fractal Concepts in Surface Growth* (Cambridge University Press, Cambridge, 1995).

[4] M.F. Shlesinger, B.J. West: Phys. Rev. Lett. **67**, 2106 (1991)

[5] B. Suki, A.-L. Barabasi, Z. Hantos, F. Peták, H.E. Stanley: Nature **368**, 615 (1994)

[6] E.R. Weibel, D.M. Gomez: Science **137**, 577 (1962)

[7] A.A. Tsonis, P.A. Tsonis: Perspectives in Biology and Medicine **30**, 355 (1987)

[8] F. Family, B.R. Masters, D.E. Platt: Physica D **38**, 98 (1989); B.R. Masters, F. Family, D.E. Platt: Biophys. J. (Suppl.) **55**, 575 (1989); B.R. Masters, D.E. Platt: Invest. Ophthalmol. Vis. Sci. (Suppl.) **30**, 391 (1989)

[9] M. Sernetz, J. Wübbeke, P. Wlczek: Physica A **191**, 13 (1992)

[10] H. Takayasu: *Fractals in the Physical Sciences* (Manchester University Press, Manchester 1990)

[11] D.R. Morse: Nature **314**, 731 (1985)

[12] T.G. Smith, W.B. Marks, G.D. Lange, W.H. Sheriff Jr., E.A. Neale: J. Neuroscience Methods **27**, 173 (1989)

[13] H.E. Stanley: Bull. Am. Phys. Soc. **34**, 716 (1989); F. Caserta, H.E. Stanley, W.D. Eldred, G. Daccord, R.E. Hausman, J. Nittmann: Phys. Rev. Lett. **64**, 95 (1990); F. Caserta, R.E. Hausman, W.D. Eldred, H.E. Stanley, C. Kimmel: Neurosci. Letters **136**, 198 (1992); F. Caserta, W.D. Eldred, E. Fernandez, R.E. Hausman, L.R. Stanford, S.V. Buldyrev, S. Schwarzer, H.E. Stanley: J. Neurosci. Methods (in press)

[14] D.R. Kayser, L.K. Aberle, R.D. Pochy, L. Lam: Physica A **191**, 17 (1992)

[15] K.R. Bhaskar, B.S. Turner, P. Garik, J.D. Bradley, R. Bansil, H.E. Stanley, J.T. LaMont: Nature **360**, 458 (1992)

[16] T. Matsuyama, M. Sugawa, Y. Nakagawa: FEMS Microb. Lett. **61**, 243 (1989); H. Fujikawa, M. Matsushita: J. Phys. Soc. Japan **58**, 387 (1989); M. Matsushita, H. Fujikawa: Physica A **168**, 498 (1990)

[17] T. Vicsek, M. Cserzö, V.K. Horváth: Physica A **167**, 315 (1990); S. Matsuura, S. Miyazima: Physica A **191**, 30 (1992)

[18] T. Vicsek: *Fractal Growth Phenomena, Second Edition* (World Scientific, Singapore 1992)

[19] E. Ben-Jacob, H. Shmueli, O. Shochet, A Tenenbaum: Physica A **187**, 378 (1992); E. Ben-Jacob, A. Tenenbaum, O. Shochet, O. Avidan: Physica A **202**, 1 (1994)

[20] H. Larralde, P. Trunfio, S. Havlin, H.E. Stanley, G.H. Weiss: Nature **355**, 423 (1992); M.F. Shlesinger: Nature **355**, 396 (1992); H. Larralde, P. Trunfio, S. Havlin, H.E. Stanley, G.H. Weiss: Phys. Rev. A **45**, 7128 (1992); S. Havlin, H. Larralde, P. Trunfio, J.E. Kiefer, H.E. Stanley, G.H. Weiss: Phys. Rev. A **46**, R-1717 (1992)

[21] J.G. Skellam: Biometrika **38**, 196 (1951)

[22] P.H. Harvey, J.R. Krebs: Science **249**, 140 (1990)

[23] R.H. Peters: *The Ecological Implications of Body Size* (Cambridge University Press, Cambridge 1983); J.E.I. Hokkanen: J. Theor. Biol. **499**, 499 (1956); A.A. Biewener: Science **250**, 1097 (1990)

[24] M. Sernetz, B. Gelleri, J. Hofmann: J. Theor. Biol. **187**, 209 (1985); R.R. Strathmann: Science **250**, 1091 (1990)

[25] J. Feder: *Fractals* (Plenum, NY, 1988)

[26] D. Stauffer, H.E. Stanley: *From Newton to Mandelbrot: A Primer in Theoretical Physics* (Springer-Verlag, Heidelberg & N.Y. 1990)

[27] E. Guyon, H.E. Stanley: *Les Formes Fractales* (Palais de la Découverte, Paris 1991); **English translation:** *Fractal Forms* (Elsevier North Holland, Amsterdam 1991)

[28] H.E. Stanley, N. Ostrowsky, eds.: *On Growth and Form: Fractal and Non-Fractal Patterns in Physics*, Proceedings 1985 Cargèse NATO ASI, Series E: Applied Sciences (Martinus Nijhoff, Dordrecht 1985)

[29] H.E. Stanley: *Introduction to Phase Transitions and Critical Phenomena* (Oxford University Press, London 1971)

[30] F. Zernike: Physica **7**, 565 (1940)

[31] H.E. Stanley, N. Ostrowsky, eds.: *Correlations and Connectivity: Geometric Aspects of Physics, Chemistry and Biology*, Proceedings 1990 Cargèse Nato ASI, Series E: Applied Sciences (Kluwer, Dordrecht 1990)

[32] P. Bak, C. Tang, K. Wiesenfeld: Phys. Rev. Lett. **59**, 381 (1987); P. Bak, C. Tang, K. Wiesenfeld: Phys. Rev. A **38**, 3645 (1988)

[33] C.-K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, F. Sciortino, M. Simons, H.E. Stanley: Nature **356**, 168 (1992)

[34] W. Li, K. Kaneko: Europhys. Lett. **17**, 655 (1992)

[35] S. Nee: Nature **357**, 450 (1992)

[36] R. Voss: Phys. Rev. Lett. **68**, 3805 (1992)

[37] J. Maddox: Nature **358**, 103 (1992)

[38] P.J. Munson, R.C. Taylor, G.S. Michaels: Nature **360**, 636 (1992)

[39] I. Amato: Science **257**, 747 (1992)

[40] V.V. Prabhu, J.-M. Claverie: Nature **357**, 782 (1992)

[41] P. Yam: Sci. Am. **267[3]**, 23 (1992)

[42] C.-K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, F. Sciortino, M. Simons, H.E. Stanley: Physica A **191**, 25 (1992)

[43] 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**, 1 (1992)

[44] C.A. Chatzidimitriou-Dreismann, D. Larhammar: Nature **361**, 212 (1993); D. Larhammar, C.A. Chatzidimitriou-Dreismann: Nucleic Acids Res. **21**, 5167 (1993)

[45] A.Yu. Grosberg, Y. Rabin, S. Havlin, A. Neer: Biofisika (Russia) **38**, 75 (1993); A.Yu. Grosberg, Y. Rabin, S. Havlin, A. Neer: Europhys. Lett. **23**, 373 (1993)

[46] J.-L. Sikorav, G.M. Church: J. Mol. Biol. **222**, 1085 (1991)

[47] A.Yu. Grosberg, A.R. Khokhlov: *Statistical Physics of Macromolecules* (Nauka Publishers, Moscow 1989)

[48] A.Yu. Grosberg, S.K. Nechaev, E.I. Shakhnovich: J. Physique **49**, 2095 (1988)

[49] K.W. Adolph, in: *Chromosome Structure and Function*, ed. by M.S. Risley (Van Nostrand, New York 1986); M. Takahashi: J. Theor. Biol. **141**, 117 (1989)

[50] S. Karlin, V. Brendel: Science **259**, 677 (1993)

[51] C.-K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, M. Simons, H.E. Stanley: Phys. Rev. E **47**, 3730 (1993)

[52] N. Shnerb, E. Eisenberg: Phys. Rev. E **49**, R1005 (1994)

[53] S.V. Buldyrev, A.L. Goldberger, S. Havlin, C.-K. Peng, M. Simons, H.E. Stanley: Phys. Rev. E **47**, 4514 (1993)

[54] A.S. Borovik, A.Yu. Grosberg, M.D. Frank-Kamenetskii: preprint

[55] S.V. Buldyrev, A.L. Goldberger, S. Havlin, C.-K. Peng, H.E. Stanley, M.H.R. Stanley, M. Simons: Biophys. J. **65**, 2673 (1993)

[56] C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger: Phys. Rev. E **49**, 1648 (1994)

[57] 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**, 64 (1994); H.E. Stanley, S.V. Buldyrev, A.L. Goldberger, S. Havlin, C.-K.Peng, M. Simons: [Proceedings of Internat'l Conf. on Condensed Matter Physics, Bar-Ilan], Physica A **200**, 4 (1993); H.E. Stanley, S.V. Buldyrev, A.L. Goldberger, S. Havlin, S.M. Ossadnik, C.-K. Peng, M. Simons: Fractals **1**, 283–301 (1993)

[58] S. Tavaré, B.W. Giddings, in: *Mathematical Methods for DNA Sequences*, Eds. M.S. Waterman (CRC Press, Boca Raton 1989), pp. 117-132

[59] J.D. Watson, M. Gilman, J. Witkowski, M. Zoller: *Recombinant DNA* (Scientific American Books, New York 1992).

[60] E.W. Montroll, M.F. Shlesinger: "The Wonderful World of Random Walks" in: *Nonequilibrium Phenomena II. From Stochastics to Hydrodynamics*, ed. by J.L. Lebowitz, E.W. Montroll (North-Holland, Amsterdam 1984), pp. 1–121

[61] G.H. Weiss: *Random Walks* (North-Holland, Amsterdam 1994)

[62] S. Havlin, R. Selinger, M. Schwartz, H.E. Stanley, A. Bunde: Phys. Rev. Lett. **61**, 1438 (1988); S. Havlin, M. Schwartz, R. Blumberg Selinger, A. Bunde, H.E. Stanley: Phys. Rev. A **40**, 1717 (1989); R.B. Selinger, S. Havlin, F. Leyvraz, M. Schwartz, H.E. Stanley: Phys. Rev. A **40**, 6755 (1989)

[63] C.-K. Peng, S. Havlin, M. Schwartz, H.E. Stanley, G.H. Weiss: Physica A **178**, 401 (1991); C.-K. Peng, S. Havlin, M. Schwartz, H.E. Stanley: Phys. Rev. A **44**, 2239 (1991)

[64] M. Araujo, S. Havlin, G.H. Weiss, H.E. Stanley: Phys. Rev. A **43**, 5207 (1991); S. Havlin, S.V. Buldyrev, H.E. Stanley, G.H. Weiss: J. Phys. A **24**, L925 (1991); S. Prakash, S. Havlin, M. Schwartz, H.E. Stanley: Phys. Rev. A **46**, R1724 (1992)

[65] C.L. Berthelsen, J.A. Glazier, M.H. Skolnick: Phys. Rev. A **45**, 8902 (1992)

[66] E.I. Shakhnovich, A.M. Gutin: Nature **346**, 773 (1990)

[67] S.G. Oliver et al.: Nature **357**, 38 (1992)

[68] E.C. Uberbacher, R.J. Mural: Proc. Natl. Acad. Sci. USA **88**, 11261 (1991)

[69] W.-H. Li, D. Graur: *Fundamentals of Molecular Evolution* (Sinauer Associates, Sunderland MA 1991)

[70] H.E. Stanley, S.V. Buldyrev, A.L. Goldberger, S. Havlin, C-K Peng, M. Simons: Physica A **200**, 4 (1993)

[71] S.V. Buldyrev, A. Goldberger, S. Havlin, C.-K. Peng, F. Sciortino, M. Simons, H.E. Stanley: Phys. Rev. Lett. **71**, 1776 (1993); R.F. Voss, Phys. Rev. Lett. **71**, 1777 (1993)

[72] W. Li: International Journal of Bifurcation and Chaos **2**, 137 (1992)

[73] J. Jurka: J. Mol. Evol. **29**, 496 (1989)

[74] R.H. Hwu, J.W. Roberts, E.H. Davidson, R.J. Britten: Proc. Natl. Acad. Sci. USA. **83**, 3875 (1986)

[75] J. Jurka, T. Walichiewicz, A. Milosavljevic: J. Mol. Evol. **35**, 286 (1992)

[76] J. Des Cloizeaux: J. Physique (Paris) **41**, 223 (1980); P.G. de Gennes: *Scaling Concepts in Polymer Physics* (Cornell Univ. Press, Ithaca 1979)

[77] G.F. Joyce: Nature **338**, 217 (1989)

[78] W. Gilbert: Nature **271**, 501 (1978)

[79] P. Hagerman: Ann. Rev. Biochem. **59**, 755 (1990)

[80] W.F. Doolitle, in: *Intervening Sequences in Evolution and Development*, ed. by E. Stone, R. Schwartz (Oxford University Press, NY 1990), p. 42

[81] A. Schenkel, J. Zhang, Y-C. Zhang: Fractals **1**, 47 (1993); M. Amit, Y. Shmerler, E. Eisenberg, M. Abraham, N. Shnerb: Fractals **2**, xxx (1994)

[82] R.N. Mantegna: Physica A **179**, 23 (1991)

[83] H.E. Stanley, N. Ostrowsky, eds.: *Random Fluctuations and Pattern Growth: Experiments and Models*, Proceedings 1988 NATO ASI, Cargèse (Kluwer Academic Publishers, Dordrecht 1988)

[84] R.I. Kitney, O. Rompelman: *The Study of Heart-Rate Variability* (Oxford University Press, London 1980); S. Akselrod, D. Gordon, F.A. Ubel, D.C. Shannon, A.C. Barger, R.J. Cohen: Science **213**, 220 (1981)

[85] M. Kobayashi, T. Musha: IEEE Trans. Biomed. Eng. **29**, 456 (1982)

[86] A.L. Goldberger, D.R. Rigney, J. Mietus, E.M. Antman, S. Greenwald: Experientia **44**, 983 (1988)

[87] J.P. Saul, P. Albrecht, D. Berger, R.J. Cohen: *Computers in Cardiology* (IEEE Computer Society Press, Washington, D.C. 1987), pp. 419–422; D.T. Kaplan, M. Talajic: Chaos **1**, 251 (1991)

[88] C.-K. Peng, J.E. Mietus, J.M. Hausdorff, S. Havlin, H.E. Stanley, A.L. Goldberger: Phys. Rev. Lett. **70**, 1343 (1993); C.K. Peng, Ph.D. Thesis, Boston University (1993); C.K. Peng, S.V. Buldyrev, J.M. Hausdorff, S. Havlin, J.E. Mietus, M. Simons, H.E. Stanley, A.L. Goldberger, in: *Fractals in Biology and Medicine*, ed. by G.A. Losa, T.F. Nonnenmacher, E.R. Weibel (Birkhauser Verlag, Boston 1994)

[89] W.B. Cannon: Physiol. Rev. **9**, 399 (1929)

[90] G.K. Zipf: *Human Behavior and the Principle of "Least Effort"* (Addison-Wesley, New York 1949)

[91] L. Brillouin: *Science and Information Theory* (Academic Press, New York 1956)

[92] C.E. Shannon: Bell Systems Tech. J. **80**, 50 (1951)

[93] L.S. Liebovitch, J. Freichbarg, J.P. Koniarek: Math. Biosci. **89**, 36 (1987); L.S. Liebovitch: Biophys. J. **55**, 373 (1989)

[94] R.V. Voss, J. Clarke: Nature **238**, 317 (1975)

[95] R.V. Voss, J. Clarke: J. Acous. Soc. Am. **63**, 258 (1978)

[96] M. Schroeder: *Fractals Chaos, Power Laws: Minutes from an Infinite Paradise* (W. H. Freeman, New York 1991)

[97] H.E. Stanley, S.V. Buldyrev, A.L. Goldberger, Z.D. Goldberger, S. Havlin, S.M. Ossadnik, C.-K. Peng, M. Simons: Physica A **205**, 214–253 (1994)

[98] N. Eldredge, S. J. Gould, in: *Models in Paleobiology*, edited by T. J. M. Schopf (Freeman and Cooper Inc., San Francisco 1972); S. J. Gould, N. Eldredge: *Nature* **366**, 223 (1993)

[99] P. Bak, K. Sneppen: Phys. Rev. Lett. **71**, 4083 (1993)

[100] S. Havlin, A.-L. Barabási, S.V. Buldyrev, C.-K. Peng, M. Schwartz, H.E. Stanley, T. Vicsek, in: *Growth Patterns in Physical Sciences and Biology* [Proc. NATO Advanced Research Workshop, Granada, Spain, October 1991], J.M. Garcia-Ruiz, E. Louis, P. Meakin, L. Sander, eds. (Plenum, New York 1993), pp. 85–98; L.-H. Tang, H. Leschhorn: Phys. Rev. A **46**, R-8309 (1992); S.V. Buldyrev, A.-L. Barabási, F. Caserta, S. Havlin, H.E. Stanley, T. Vicsek: Phys. Rev. A **46**, R-8313 (1992); A.-L. Barabási, S. V. Buldyrev, S. Havlin, G. Huber, H. E. Stanley, T. Vicsek, in: R. Jullien, J. Kertész, P. Meakin, D. E. Wolf (eds.), *Surface Disordering: Growth, Roughening, and Phase Transitions* [Proc. of the Les Houches Workshop 1992] (Nova Science, New York 1992), pp. 193–204