ELSEVIER

Physica A 224 (1996) 302-321

PHYSICA A

Anomalous fluctuations in the dynamics of complex systems: from DNA and physiology to econophysics

H.E. Stanley[a], V. Afanasyev[b], L.A.N. Amaral[a], S.V. Buldyrev[a], A.L. Goldberger[c,d], S. Havlin[a,e], H. Leschhorn[a], P. Maass[a], R.N. Mantegna[f], C.-K. Peng[a,c], P.A. Prince[b], M.A. Salinger[g], M.H.R. Stanley[a], G.M. Viswanathan[a]

[a] *Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215, USA*
[b] *British Antarctic Survey, Natural Environment Research Council, High Cross, Madingley Road, CB3 0ET, Cambridge, UK*
[c] *Cardiovascular Division, Harvard Medical School, Beth Israel Hospital, Boston, MA 02215, USA*
[d] *Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA*
[e] *Minerva Center and Dept. of Physics, Bar-Ilan University, Ramat Gan, Israel*
[f] *Dipartimento di Energetica ed Applicazioni di Fisica, Palermo U., Palermo, I-90128, Italy*
[g] *School of Management, Boston University, Boston, MA 02215, USA*

Abstract

We discuss examples of complex systems composed of many interacting subsystems. We focus on those systems displaying nontrivial long-range correlations. These include the one-dimensional sequence of base pairs in DNA, the sequence of flight times of the large seabird *Wandering Albatross*, and the annual fluctuations in the growth rate of business firms. We review formal analogies in the models that describe the observed long-range correlations, and conclude by discussing the possibility that behavior of large numbers of humans (as measured, e.g., by economic indices) might conform to analogs of the scaling laws that have proved useful in describing systems composed of large numbers of inanimate objects.

1. Introduction

In recent years anomalous fluctuations have been discovered in a remarkably wide variety of phenomena from DNA sequences [1] and heartbeat intervals [2] to complex behavior of animals [3] and even to behavior of economical systems such as stock market fluctuations [4] or fluctuations of firm sales [5,6]. The existence of anomalous

Elsevier Science B.V.
SSDI 0378-4371(95)00409-2

fluctuations may indicate analogies in the underlying mechanisms in totally different systems which from the first glance do not have anything in common with each other [7].

We will concentrate on two such analogies:

1. Long-range correlations in DNA sequences and anomalous power law dependence of sales fluctuations of industrial firms on their size.
2. Patterns in DNA sequences and patterns of sea-bird behavior.

The paper is organized as follows. In Section 2 we discuss the general concept of long-range correlations in DNA sequences, In Section 3 we discuss the duplication-deletion model of DNA evolution related to the model of hierarchical firm structure, discussed in Section 4. In the final two sections we discuss the Lévy-walk models of long range correlations in DNA (Section 5) and sea-bird foraging (Section 6). Finally, we conclude with some discussion of the possibility that behavior of large numbers of humans (as measured, e.g., by economic indices) might conform to (or be usefully understood using) analogs of the scaling laws that have proved useful in describing systems composed of large numbers of inanimate objects.

2. Long-range correlations in DNA sequences

In order to study the scale-invariant long-range correlations of a DNA sequence, we first introduce a graphical representation of DNA sequences, which we term a *fractal landscape* or *DNA walk* (Fig. 1) [1]. For the conventional one-dimensional random walk model [8,9], 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 [10–12].

There are actually many possible rules of mapping of DNA sequence onto one-dimensional random walk [11,13]:

(i) $u(i) = +1$ for C or T and $u(i) = -1$ otherwise ("purine-pyrimidine" rule);

(ii) $u(i) = +1$ for C or G and $u(i) = -1$ otherwise; ("hydrogen bond" rule)

and so on, see Ref. [14] and references therein. The energy rule as well as the analogy of the DNA sequences and the Ising system was proposed, and the existence of large domains of CG rich versus AT rich regions was observed (see, e.g., Ref. [13]). 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).

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. Landscapes for intron-containing

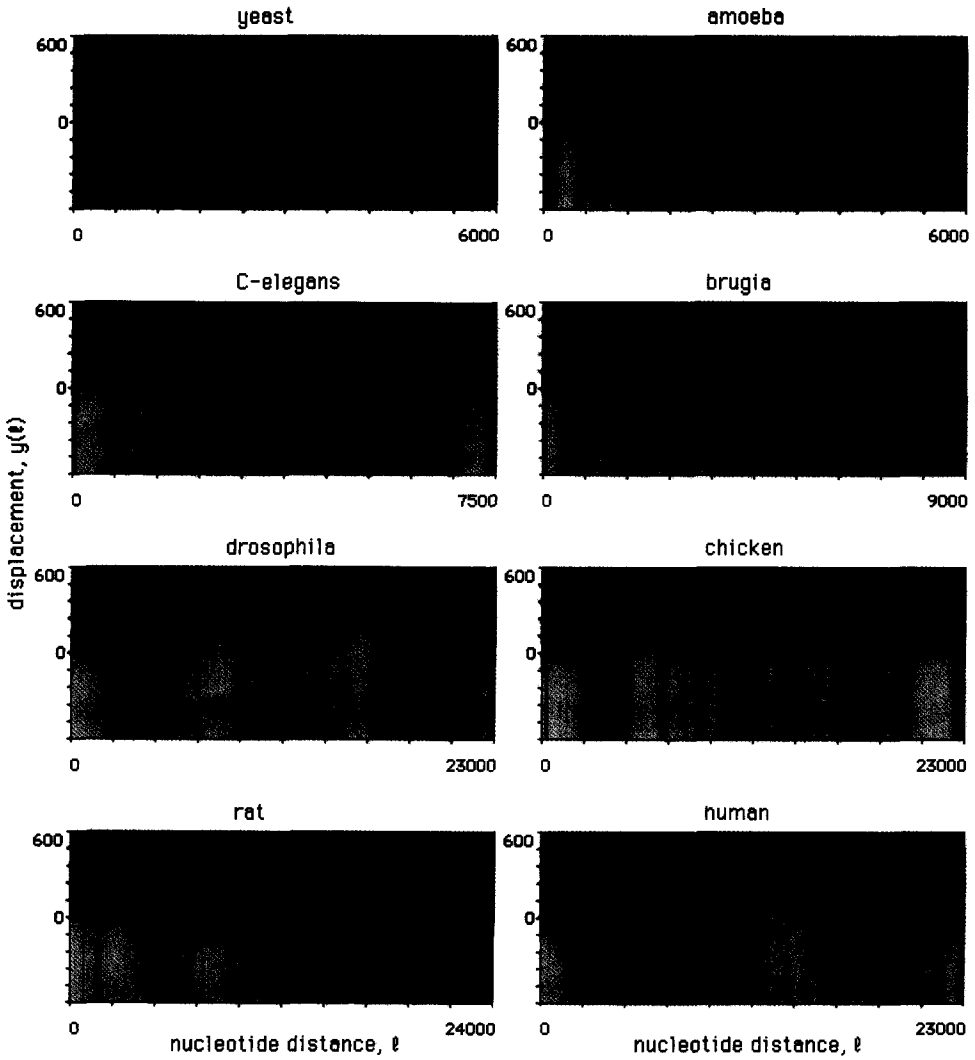


Fig. 1. The DNA walk representations of 8 DNA sequences from the MHC family. DNA landscapes are plotted so that the end points have the same vertical displacement as the starting points. The graphs are for yeast, amoeba, worms: *C. elegans*, *Brugia malayi*, drosophila, chicken, rat and human (from top to bottom, left to right). The dark areas denote coding regions of the genes. The DNA walks for the genes show increasing "complexity" with evolutions. After [33].

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. Such DNA landscapes naturally motivate 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 $y(\ell) \equiv \sum_{i=1}^{\ell} u(i)$.

An important statistical quantity characterizing any walk [8,9] 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,

$$F^2(\ell) \equiv \overline{[\Delta y(\ell) - \overline{\Delta y(\ell)}]^2} = \overline{[\Delta y(\ell)]^2} - \overline{\Delta y(\ell)}^2, \tag{1}$$

of a quantity $\Delta y(\ell)$ defined by $\Delta y(\ell) \equiv y(\ell_0 + \ell) - y(\ell_0)$. 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 $\ell_0 = 1$ to $\ell_0 = 2, \cdots$ and (c) calculating the quantity $\Delta y(\ell)$ (and its square) for each value of ℓ_0, and (d) averaging all of the calculated quantities to obtain $F^2(\ell)$.

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

$$C(\ell) \equiv \overline{u(\ell_0)u(\ell_0 + \ell)} - \overline{u(\ell_0)}^2, \tag{2}$$

through the relation

$$F^2(\ell) = \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} C(j - i). \tag{3}$$

A different way [14–16] of quantifying the fluctuations in DNA sequences is to compute power spectrum $S(f)$ of the sequence $u(i)$, which is obtained by (a) Fourier transforming the sequence $\{u(i)\}$ and (b) taking the squared absolute value of complex Fourier component with frequency f. For a stationary sequence, the power spectrum is the Fourier transform of the correlation function.

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 $\alpha \neq 1/2$.

In the last case we expect power-law behavior for both the power spectrum and the correlation function,

$$S(f) \sim (1/f)^{\beta}, \tag{5}$$

and

$$C(\ell) \sim (1/\ell)^{\gamma}. \tag{6}$$

The correlation exponents α, β and γ are not independent, since [10,11]

$$\alpha = \frac{1+\beta}{2} = \frac{2-\gamma}{2}. \tag{7}$$

For non-stationary, "patchy", sequences, such as coding DNA sequences the behavior of $F(\ell)$ 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 [1,17,18].

On the other hand, the data for intron-containing and intergenic (i.e. noncoding) sequences are more linear on this double logarithmic plot and can be well approximated by 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.

In order to be able to correctly quantify the fluctuations in patchy sequences, without splitting them into separate patches, we develop the detrended fluctuation analysis (DFA) described in detail in Refs. [14,19] which is based on the computation of the detrended fluctuation $F_d(\ell)$, and takes into account the local trend in each patch (Fig. 2). Thus DFA is designed to treat sequences with statistical heterogeneity such as DNA's known mosaic structure ("patchiness") arising from non-stationarity of nucleotide concentration. On the other hand, for stationary sequences, such as inverse Fourier transform of a given power spectrum with random phases [11,20,21], DFA produces exactly the same results as original fluctuation function analysis. The advantage of the DFA method is that the $F_d(\ell)$ function for a patchy landscape that consists of alternating uncorrelated patches of fixed length ξ has a slope $\alpha = 1/2$ for all length scales $\ell < \xi$ and then indicates a sharp maximum in α around $\ell = \xi$. Thus DFA is able to detect characteristic length scales the same way the standard power spectrum technique does, but it is more accurate due to the fact that $F^2(\ell)$ is a double summation of $C(\ell)$. Thus it would seem that the DFA is more useful than the power spectrum method due to reduced noise. For a systematic analysis of the finite size effects on the exponent α, see [22] and for applications of long-range correlations see [23].

The articles by Peng [1], Li [15,24], Voss [16], Karlin and Brendel [17], and Arneodo [25] have raised a question in computational molecular biology whether long-range correlations are present in both coding and noncoding DNA or only in the latter. To answer this question, we consider [14] 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. Standard power specrum analysis Fig. 3a) indicates that *coding* sequences have practically no correlations in the range from 10 bp to 100 bp (spectral exponent $\beta \pm 2SD = 0.00 \pm 0.04$ or $\alpha = 0.50 \pm 0.02$). In contrast, for *noncoding* sequences, the average value of the spectral exponent β is positive ($\beta = 0.16 \pm 0.05$ or $\alpha = 0.58 \pm 0.03$), which unambiguously shows the presence of long-range correlations. We also separately analyze the 874 coding and 1157

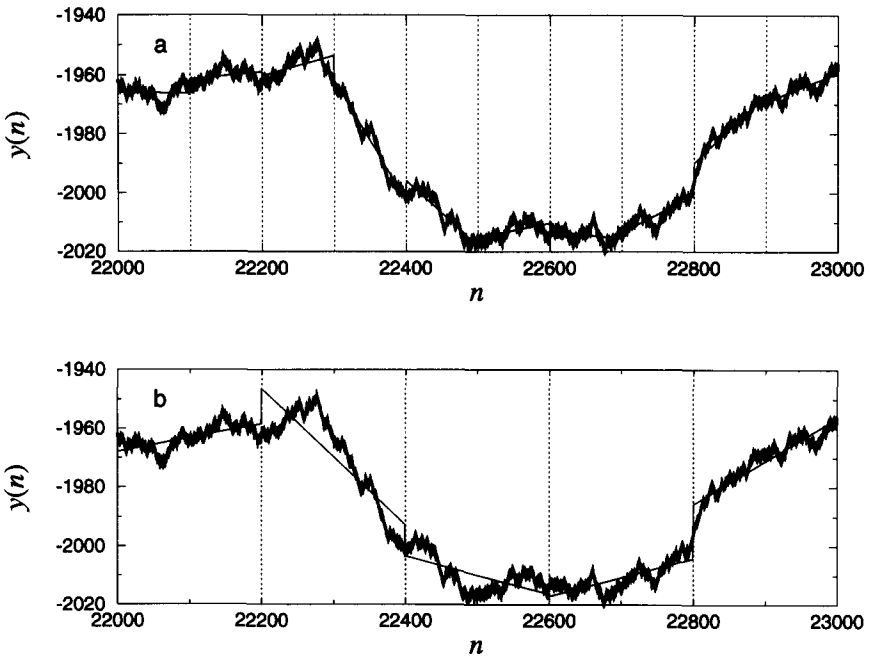
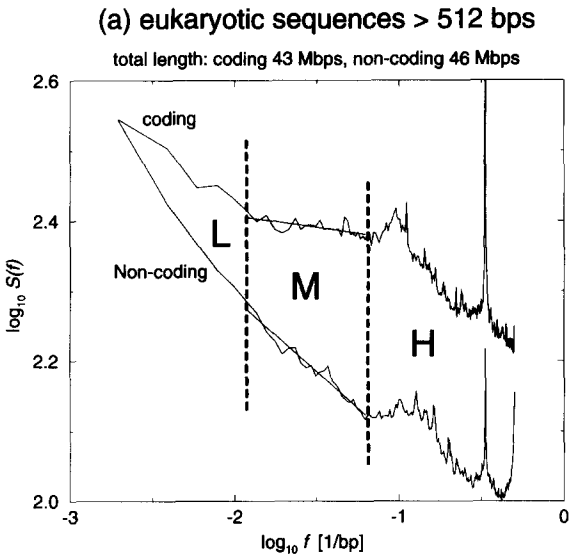


Fig. 2. "DNA walk" generated by a subsequence of the bacteriophage λ genome. The detrended fluctuation analysis (DFA) is applied in (a) to box size $\ell = 100$, and in (b) to box size $\ell = 200$. Shown in each box is the least squares fit to the data in that box. Note that the typical variance for a box in (b) is larger than for a box in (a). The quantitative fashion whereby the variance increases with box size determines the long-range correlation exponent α.

noncoding sequences which have more than 4096 bp (Fig. 3b), and find a larger region of power law behavior. We calculate the probability that these two data sets (coding and non-coding) were drawn from the same distribution, and we find that it is less than 10^{-10}. We obtain independent confirmation of these findings using the method of detrended fluctuation analysis. The near-perfect agreement between the two independent analysis methods, FFT and DFA, increases the confidence in the reliability of our conclusion. Thus the coding sequences are practically uncorrelated up to the length of coding sequence of single protein, which rarely exceeds several thousands nucleotides, while non-coding sequences and entire chromosomal regions including both coding and non-coding sequences have long-range correlations of all length scales.

3. Duplication-mutation model of DNA evolution

The question arises whether these long range correlations in non-coding DNA sequences and the entire chromosomes are the simple consequence of patches of DNA with different nucleotide concentration [17]. Indeed, how can power law correlations arise in the one-dimensional system such as DNA, where correlations should decay exponentially with distance between nucleotides in analogy with spins of one-dimensional



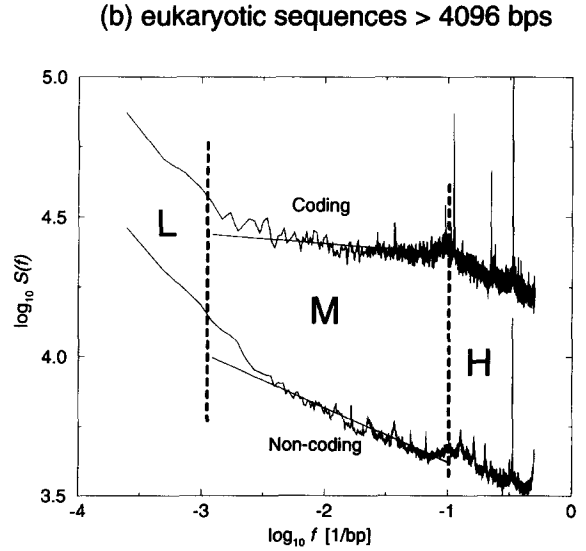


Fig. 3. (a) Power spectra averaged over all eukaryotic sequences longer than 512 bp, obtained by FFT with window size 512. Upper curve is average over 29 453 coding sequences; lower curve is average over 33 301 noncoding sequences. The straight lines are least squares fits for second decade (Region M). The values of β measured as the slopes of the fits are 0.03 and 0.21, respectively. (b) Same data for all sequences larger than 4096 bp, obtained by FFT with window size of 4096. The average is computed over 874 coding and 1157 noncoding sequences. Note that for high frequencies, the power spectra for both window sizes practically coincide. In the region of frequencies $f < 1/100\ \mathrm{bp}^{-1}$ (Region H on Fig. 3a), the power spectra in (a) bend upward from the apparent straight line. For (b) (larger windows) the $S(f)$ spectra have constant slope over more than one decade (region M). The fits are the same for both (a) and (b): for coding, $\beta = 0.04$, while for noncoding, $\beta = 0.21$.

Ising models? One of the possible answers to this question is the duplication-mutation model of DNA evolution suggested by Li [26]. (See also Bell [27] and Sutherland and Richards [28] for other hypotheses.)

In this model the time axes serves as an additional spatial dimension which connects distant segments of DNA which have been developed from a single ancestor.

The model is based on two assumptions both of which are well biologically motivated:

1. Every nucleotide can mutate with certain probability.
2. Every nucleotide can be duplicated or deleted with certain probability.

First phenomenon is known as point mutation which can be caused by random chemical reactions such as methylation [29]. Second phenomenon often happens in the process of cell division (mitosis and myosis) when pairs of sister chromosomes exchange segments of their DNA (genetic crossover). If the exchanging segments are of identical length the duplication does not happen. However, if two segments differ in length by n nucleotides, the chromosome that acquires larger segment obtains an extra sequence of length n which is identical to its neighbor , while another chromosome loses this sequence. Thus a tandem repeat of length n appears on one of the sister chromosomes. In many cases duplications can be more evolutionary advantageous than deletions. In this case lengthy tandemly repeated regions will emerge from a single repeat. For simplicity we will start with a model similar to the original model of Li [26] which neglects deletions and deals with duplication of single nucleotides ($n = 1$). Next we will discuss the implications of deletions.

Schematically, this model can be illustrated by Fig. 4a. Each level of the tree-like structure represents one step of evolution process during which each nucleotide always duplicates and with probability p_m also mutates. For simplicity we assume only two types of nucleotides a and b (say purine vs. pyrimidine) each of which is represented by a step up or down in the DNA walk representation. After k steps, this process will lead to a sequence of 2^k nucleotides which is represented by DNA landscape, shown below. The total excess of purines over pyrimidines

$$\Delta y = \sum_{i=1}^{2^k} u_i, \tag{8}$$

is equal to the difference in heights of the starting and the ending points of the walk. In the following we compute explicitly the correlation

$$C(\ell) = \langle u_i u_{i+\ell} \rangle \tag{9}$$

between nucleotides which are ℓ nucleotides apart from each other along the resulting sequence. The reason of why the correlations are now long-range is obvious. Indeed, the nucleotides which are $\ell = 2^{k'}$ apart from each other in space are only $2k'$ apart from each other in time, since they are both descendants of one common ancestor $k' = \log_2 \ell$ generations before. The correlation decay exponentially with k' and hence as a power law of ℓ.

Simple calculations yield

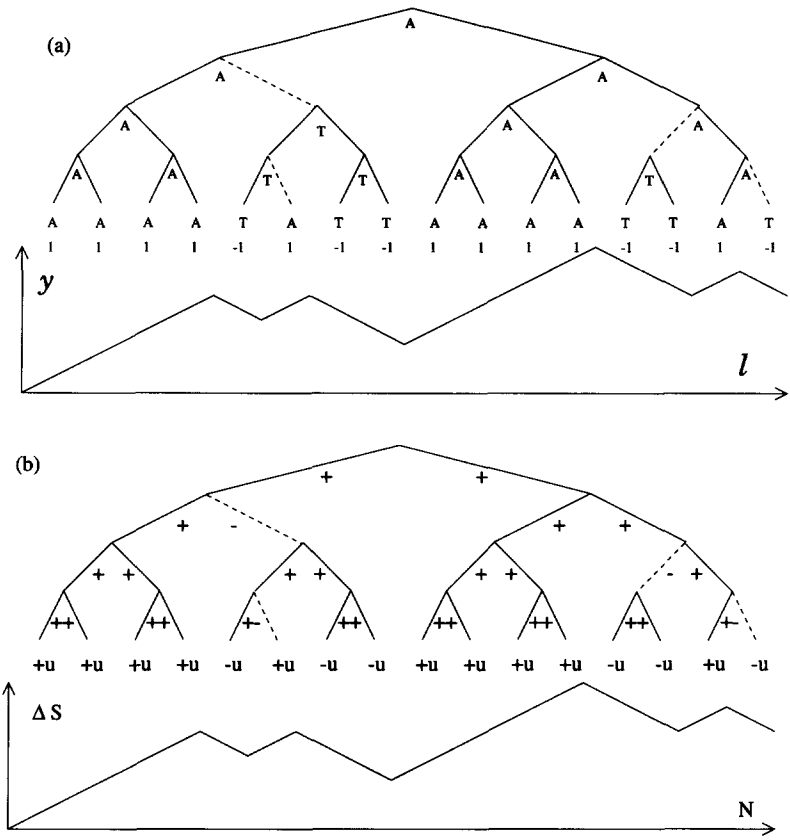


Fig. 4. (a) Schematic representation of the most simple example of duplication-mutation process proposed by W. Li. Originally the sequence consists of only one nucleotide A. At any time step, represented by a certain horizontal level of the tree each nucleotide duplicates and some of them mutate with probability $p_m \ll 1$. Mutation events are shown by dashed lines. For simplicity we assume only two nucleotides A and T. The DNA walk representation of the obtained sequence is shown below the tree. (b) Analogous plot for the structure of a firm. Dashed lines correspond to modifications of the "bosses" decisions by lower level management. To total change in sales can be calculated by adding changes u_i of each branch of a firm listed below the tree.

$$\langle u_i u_{i+\ell} \rangle = (1 - 2p_m)^{2k'} = \ell^{\frac{2\ln(1-2p_m)}{\ln 2}}. \tag{10}$$

Thus

$$\gamma = -\frac{2\ln(1-2p_m)}{\ln 2}, \tag{11}$$

and, using (7),

$$\alpha = 1 - \frac{|\ln(1-2p_m)|}{\ln 2}. \tag{12}$$

Note that $\alpha = 1$ when $p_m = 0$ and α becomes $1/2$ when $p_m > \frac{1}{2}(1 - \frac{1}{\sqrt{2}})$.

In general, when the deletions might occur with some probability $p_d < 1/2$, the number of descendants of one common ancestor grows as $z^{k'}$ where $z = 2(1 - p_d)$ and k' is the number of generations.

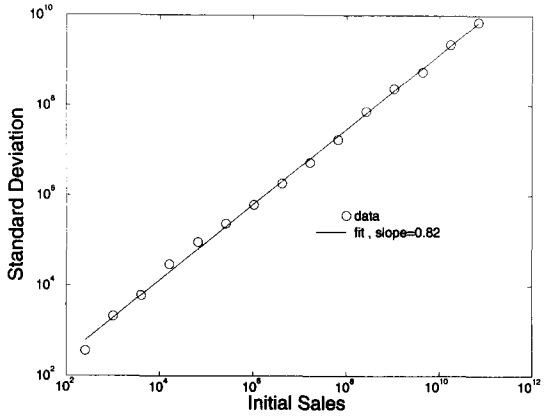


Fig. 5. Power-law dependence of standard deviation $\sigma(S)$ on initial sales S.

Thus, replacing $\ln 2$ by $\ln z$ in the denominators of expressions (11) and (12), we get

$$\gamma = \frac{2\ln(1-2p_m)}{\ln(2-2p_d)}, \qquad \alpha = 1 - \frac{2\ln(1-2p_m)}{\ln(2-2p_d)}. \tag{13}$$

More rigorous but less evident approach of recursion relations among levels of the tree lead to the same analytical results - see Eqs. (11), (12), and (13).

Similar arguments can be applied for computation of α in more complex situations when more then one nucleotide can duplicate and all four types of nucleotides are present, however simple analytical results in this case are not available.

In summary the model suggested by Li may lead under reasonable assumptions to the experimentally observed values of α which are in the range between 0.5 and 1. In the next section, we show how this model can be applied to the study of an economic system.

4. Annual fluctuations of firm growth rate

Another quite unrelated phenomenon is the behavior of industrial firm sales or their employment. We have studied the dependence of the fluctuations of the annual firm growth rates on the initial size of the firm [6]. We computed the average annual fluctuation of employee numbers or sales $\sigma(S_0)$ as a function of the initial value of sales or employees S_0. The remarkable linearity of the $\sigma(S_0)$ vs. S_0 function on a log-log scale over many orders of magnitude may indicate some universal law of economics that is applicable for small companies with sales of several thousands dollars per year as well as for giants of size of General Motors with hundred billion annual sales. We found that $\sigma(S_0) \sim S_0^{\alpha}$, where $\alpha = 0.82 \pm 0.05$ (Fig. 5).

The power law increase of $\sigma(S_0)$ on S_0 may have its origin in the internal structure of each firm. In the simplest approach, one would assume that the sales S_0 of a given

company result from N independent units

$$S_0 = \sum_{i=1}^{N} \xi_i. \tag{14}$$

If the unit sales ξ_i have a typical average $\langle \xi \rangle = S_0/N$ and an annual variation u_i independent of s_0. Then the annual change in sales is

$$\Delta S = \sum_{i=1}^{N} u_i. \tag{15}$$

In analogy with a random walk, $\sigma(S_0)$ would grow as $\sqrt{N}$ or since N is proportional to S_0 as $\sqrt{S_0}$, thus giving $\alpha = 1/2$. The much larger value of α that we find indicates the presence of strong correlations among the firm's units. We can model this phenomena by considering the tree-like hierarchical organization of a typical firm (see Fig. 4b). The root of the tree represents the head of the company, whose policy is passed to the level beneath, and so on, until finally the units in the lowest level take action. Each of these units has an average sales value $\langle \xi \rangle = S_0/N$ and a corresponding typical fluctuation u. The number of links connecting the levels will vary from level to level, but there is a value z which represents a certain average number of links. Then the number of units N is equal to z^k, where k is the number of levels.

What are the consequences of this simple model? Let us first assume that the head of the firm suggests a policy that could result in changing the sales of each unit by an amount u. If this policy is propagated through the hierarchy without any modifications, then the change in sales is simply $\Delta S = N_u = S_0 u/\langle \xi \rangle$. Accordingly, $\alpha = 1$.

More realistically, the policy of the head can be modified (undergo "mutation") at each level of the firm management with a small probability p_m. Hence the sales of the entire firm becomes a random variable with a standard deviation that can be explicitly computed using recursion relations among the levels of the tree. The result is $\sigma^2(S_0) = u^2[4zp_m(1-p_m)(y^k - z^k)/(y-z) + y^n]$, where $y = z^2(1-2p_m)^2$ and $k = \ln(S_0/\langle \xi \rangle)/\ln z$. For large k, the model predicts $\alpha = 1 - |\ln(1-2p_m)|/\ln(z)$ if $z > 1/(1-2p_m)^2$ and $\alpha = 1/2$ otherwise [see Eq. (12)]. Despite the model's simplicity, this result seems to be plausible. For example, if we require $\alpha = 0.8$, as suggested by our data, and choose $z = 3$, we obtain $p_m = 0.1$, which might be a reasonable situation.

The models proposed are very elementary, and show that simple mechanisms can provide some insight into our findings.

Remarkably, the hierarchical structure of the company (Fig. 4b) can be mapped exactly onto the diagram of the DNA mutations and duplications (Fig. 4a). Each level of the firm hierarchy corresponds to one generation of repeat family and each modification of the head decision by the lower level management corresponds to a mutation. Note that the $\sigma(S_0)$ for firm sales is exactly $F(\ell)$ for DNA sequences.

5. Lévy walk model of DNA sequences

Apparently, the molecular evolution of DNA is too complicated process to be explained by a single model. The duplication-mutation model described in Section 3 can be used to explain certain features of highly repetitive DNA on relatively short length scales (such as tandem repeats) and probably explains the difference of statistical properties of coding and noncoding DNA. In coding DNA, where duplications and mutations would in most of cases lead to fatal consequences, the long-range correlations cannot develop. However this model does not take into account many other important processes of DNA evolution like retroviral insertions and deletions, which are probably the main source of the rapid evolution of DNA sequences on the large length scales, comparable to the length of the entire chromosome. On such length scales the question of the difference between coding and non-coding DNA is no longer valid since the coding sequences are limited to a protein size.

An example of such retroviral insertions is the LINE-1 sequence which consists of 6,139 base pairs and is believed to contain a code for a functional protein [30]. In agreement with this it is found that the LINE-1 sequence has value of α close to 0.5, indicating the lack of long-range correlations [18]. Moreover, the LINE-1 sequence has a strong strand bias of about 59% of purines, which is also typical for coding sequences. The total number of LINE-1 sequences and fragments in the human genome is estimated to be 107,000, while in the genome of the chimpanzee there are only 51,000 copies of the LINE-1 sequence [31]. This dramatic difference indicates that thousands of insertions or deletions of LINE-1 sequences took place over a relatively short evolutionary time scale. LINE-1 sequences are found on both strands of DNA and therefore produce large local fluctuations of nucleotide content. Another frequent repetitive element is the ALU sequence [32], which is also statistically similar to protein coding DNA, but, in contrast with the LINE-1 sequence, is only 290 base pairs long.

The central idea of the insertion model [18,33] is based on the assumption that the insertion of retroelements, formed by the inverse-transcribed RNA, plays a major role in DNA evolution (see also Ref. [29]). The statistical properties of retroelements are similar to those of protein coding sequences. In order to be inserted into DNA, a retroelement must form a loop. The probability to find a loop of certain size ℓ in a long polymer chain in a solvent is given [34] by the formula

$$P(\ell) \propto (1/\ell)^{\mu}, \tag{16}$$

where μ is a critical exponent with a value close to 2.2. Thus we assume

(i) that DNA sequences are comprised of subsequences distributed according to Eq. (16), and

(ii) that these subsequences are statistically similar to protein coding sequences which (a) usually have a significant excess of purines over pyrimidines (or vice versa because of DNA two-strand complementarity) and (b) can be modeled by a Markovian process with short range correlations [35].

This biological evolution model is mathematically equivalent to the generalized Lévy walk which gives rise to a landscape which consists of alternating patches of different orientation, whose length distribution obeys power-law given by Eq. (16). The Lévy-walk model has a well defined power-law long-range correlation exponent α that depends upon the Lévy walk parameter μ [18,12]

$$\alpha = \begin{cases} 1, & \mu \leq 2, \\ 2 - \mu/2, & 2 < \mu < 3, \\ 1/2, & \mu \geq 3. \end{cases} \tag{17}$$

Thus nontrivial behavior of α corresponds to the case $2 < \mu < 3$, where the first moment of $P(\ell)$ converges while the second moment diverges. The long-range correlation property for the Lévy walk, in this case, is related to the broad distribution of Eq. (16) that lacks a characteristic length scale. Eq. (17) is valid only asymptotically for large values of ℓ. For small ℓ the slope of the log-log plot of the function $F(\ell)$ for the generalized Lévy walk model increases monotonically from a value defined by short range Markovian correlations of the inserted subsequences to a value $\alpha = 0.9$ predicted by Eq. (17). However, this limiting value can be achieved only for very long sequences of about 10^6 base pairs, and has a large standard error for finite sequences [22].

To test the insertion model [18], we have adjusted its parameters, to best approximate features of actual DNA sequences and generate an artificial sequence that corresponds to Eq. (16) and found a good agreement between the model and the actual data on successive slopes of the $F(\ell)$ function for all sequences, that contain a substantial percentage of noncoding material.

6. Lévy-walk model of sea bird foraging

Similar patchy behavior have been recently observed in the foraging pattern of the Wandering Albatross [36]. Recently, several reports have raised the possibility that some biological systems have scale invariant properties [37–46]. However the basis for such scale invariant behavior has remained elusive. In case of the Wandering Albatross *Diomedea exulans*, the origin of this scale invariant temporal behavior may be related to the spatial scale invariance in the underlying ecosystem [47].

Temporal behavior of sea birds can be easily quantified: the sea birds divide their time entirely between either flying or sitting on the water surface. An electronic recording devices was attached to the legs of 19 birds that took measurements every 3 s, and recorded $u(t)$, the number of 15 s intervals in each hour for which the animal was wet for 9 s or more [36]. Each entry in the time series $u(t)$ is therefore a number from 0 to 240, and $t = 1, 2, \ldots t_{\max}$ is time measured in hours. For the 19 sea birds studied, 19 such time series were obtained, the shortest having $t_{\max} = 77$ h and the longest $t_{\max} = 416$ h (the time series have an average length of 175 h). The obtained time series can be mapped to one-dimensional walk where the difference in heights of two points is equal to

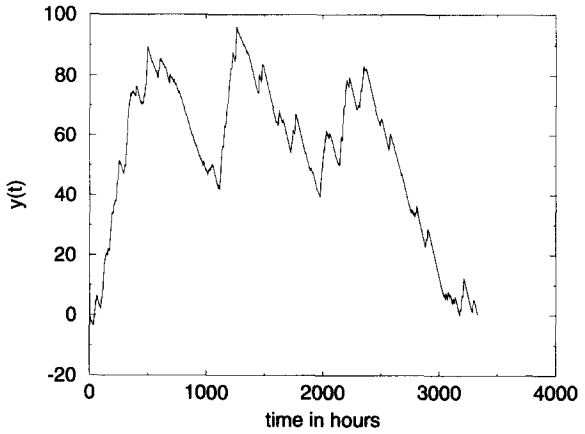


Fig. 6. The one-dimensional walk landscape, constructed from all 19 time series of bird foraging [36] analyzed using Eq. (18). Note the remarkable visual similarity with DNA landscapes (see Fig. 1).

$$\Delta y = \sum_{i=t_0}^{t_0+t} u(i) - t\langle u(i)\rangle, \tag{18}$$

see Fig. 6. The average value of the time spent on water $\langle u(i)\rangle$ is subtracted for clarity. The obtained landscape is very similar to those of long DNA sequences, accordingly, we applied the same methods for measuring long-range correlation exponents. Here we obtained $\alpha \approx \beta \approx 1.0$, using both power spectrum analysis and DFA. We also measured the distribution of uninterrupted flight times T_i which correspond to the patches with large negative slope on Fig. 6 and find that the probability to find a flight with length T decays as $T^{-\mu}$ with $\mu \approx 2$ which is in a good agreement with Lévy walk theory [see Eq. (17)]. The power law distribution of flight times is consistent with the hypothesis of fractal distribution of the food (plankton) on the surface of the ocean. Indeed, suppose that the bird is selecting the direction of each flight at random and that it flies with constant speed which does not depend on the direction of flight. Then each flight will correspond to a segment of a straight line that connects two consecutive landing points. The distribution of these segments will again follow a power law with the same exponent μ. This pattern exactly corresponds to the Lévy-flight model described in Ref. [48]. The fractal dimension of the set of landing points of a Lévy-flight is $D = \mu - 1$ (for more details, see Ref. [49]). In our particular case we get fractal dimension close to 1. Assuming that the density of landing points is proportional to the density of food, we conclude the plankton forms a fractal set on the surface of the ocean.

7. Outlook

One theme of this talk has been that the ideas of phase transitions and critical phenomena may have utility in explaining complex behavior of living systems – up to and including the science of human behavior. It is difficult to obtain large databases on

human behavior unless we turn to economics, where not only does a wealth of data exist but also the "human behavior" is subject to well-defined "rules." Like a game of chess, the goal is clear (the analog of catching the king is making a profit). Moreover, many things that would help reach the goal are illegal (just as the chess players cannot move freely on the chessboard, so also those who play the game of business are constrained in their actions by a well-defined set of rules).

So far, we have discovered the intriguing result that there is an exponent that governs some economic behavior over roughly 8 decades. We conclude by discussing briefly the implications of this result. A question asked often of us is why the result of a power law with exponent different from 1/2 is not merely what anyone would expect. Indeed, correlations of necessity exist, and therefore the exponent should change from the value 1/2 expected for the uncorrelated case. Moreover, the very question of the existence of an exponent at all has *not* been posed previously.

Some historical perspective may be illuminating at this point. In the early 1960's, exponents were as new to critical point phenomena as they are now to economics. Whenever someone would give a talk on exponents, several in the audience would make uncomplimentary remarks to the effect: "of course without correlations, you must get the mean field exponent of 1/2, but when there are correlations you will get a different exponent – so what?" Then the speaker would need to patiently explain that the important thing is the nature of the correlations could be *quantified* using the value of the exponent. An exponent close to the mean field value might indicate in some sense "less departure" from the uncorrelated case than an exponent far from the mean field value. And the speaker might explain that most people think of fluctuations as uncorrelated, but the numerical value of an exponent provides a quantification of the correlations in this randomness. Armed with this quantification, it became possible to make a serious comparison of theory and experiment, and it became possible to compare experiments on different systems. We thereby learned which features of a system are important for determining the exponent (i.e., the nature of the random fluctuations) and which features are unimportant. The analogy between economics and critical phenomena is sufficiently strong that a similar story might evolve.

A second concern is as follows. Exponents in critical phenomena are associated with critical points. But the economy is not at a critical point all the time – so what conceivable use could be the analogy of critical point exponents and the exponents characterizing the set of 4000 US publicly-traded firms? Of course the economy is not perpetually at a "critical point" of the conventional sort. However in recent years more and more out-of-equilibrium or "dynamical" systems have been found to be describable using the same conceptual framework as that discovered 30 years ago to describe equilibrium critical points. Such out-of-equilibrium systems include many forms of disorderly growth, such as diffusion limited aggregation – where the rules are simple but the resulting object is quite disorderly and is certainly not in equilibrium.

Some problems can be viewed as either equilibrium or out-of-equilibrium. An example is percolation. We can make an ensemble of "equilibrium" percolation clusters by assigning random numbers to all the plaquettes of a lattice, and coloring black

those whose random number is smaller than 0.593. Alternatively, we can create an out-of-equilibrium "dynamical" version of percolation – termed invasion percolation – by coloring black any randomly chosen site, and then iterating the simple rule that we blacken the perimeter site characterized by the smallest random number. One thereby grows a single large cluster that is automatically "at" the critical point in that as it grows larger and larger, it becomes closer and closer in every measurable property to the incipient infinite cluster that is found in conventional percolation.

The set of such systems that are clearly out-of-equilibrium and that evolve dynamically in time has expanded considerably since DLA and invasion percolation. It is worth studying the degree to which the economy is a representative of such systems. To this end, an important first step is to discover whether quantifiable features of the economy possess the hallmarks of economic systems, such as self-similarity (i.e., scaling exponents). That we have identified one quantifiable feature (the growth rate) and discovered empirically that such a scaling law exists fits therefore the tradition of scientific enquiry.

Finally, one wonders if what we discovered is so straightforward, why was it not done before? There are two answers to this question.

1. In order to make the discovery of scaling in the growth rate distribution, a truly gargantuan amount of data analysis was required. Specifically, it was necessary to measure properties (sales and employment) of every US publicly-traded firm each year for a 20-year period. The entire analysis of growth rate statistics was repeated for employment and assets, for confirmation of the general principle.
2. The second answer is that the theoretical economics community proceeds by a somewhat different route than the theoretical statistical physics community proceeds. Economists frequently make a theory or model, and later test the model using real data. Statistical physicists, on the other hand, may instead study empirical data hoping to find a pattern, trend, or "scaling law" and only much later (if ever) explain the scaling law. Indeed, our approach is called "empirical" by the economists. A second difference is that economists rarely use graphical representations of the data as extensively as we do. An advantage of our approach is that we can discover "patterns" (such as the scaling laws); indeed, the likelihood of predicting these patterns in advance of the sort of graphical data analysis we undertook is probably fairly low.

Finally, we address the potential utility of the discovery we report. There are often two different answers to questions regarding utility.

1. The first answer concerns "scientific" utility – utility in pushing back the frontiers of scientific understanding. The theory of the firm is a fundamental problem in economics that remains unsolved and that has been a source of considerable interest in recent years. The discovery of "scaling invariance" over a range of 10^7 means that a tiny firm with sales of only 10^4 dollars somehow obeys the same law as a huge firm with sales of 10^{11} dollars. One mindset in the economics community has involved partitioning the entire economy into different sectors. Our work suggests that all the firms interact with all the other firms. That the economy consists of many many strongly interacting firms, just as a fluid near its critical point consists of many strongly-interacting

particles. One cannot say much about the nature of the interactions of the firms, just as one cannot say much about the nature of the interactions of the particles in a critical system. Nonetheless, the quantitative characterization of both firms and critical systems can be carried out in terms of scaling laws. These are "laws of economics" in the same sense that the scaling laws in physics are "laws of physics": they are valid, even though a completely firm theoretical foundation is lacking.

2. The "practical" utility of our work concerns the predictive ability of the scaling laws. Specifically, from these scaling laws one can use the 1995 sales data to predict a set of histograms for the year 1996. That is, given the firm sales in 1995, one can predict the histogram of firm sales in 1996. We can now make quantitative predictions by actually giving the complete probability distribution and the quantitative parameters that characterize this probability distribution.

There are, of course, things one cannot predict:

1. It is important to emphasize, before concluding, that we cannot predict how any given firm will do next year, any more than the Gutenberg-Richter plot of earthquake frequency against earthquake magnitude can predict when the next earthquake in London will occur.
2. It is also important to emphasize that while we can predict the width (or standard deviation) of the histograms giving the number of firms with sales of X, we cannot predict the "mean." The analog of this statement in turbulence [50] is that one cannot predict the actual velocity of a red cork tossed into a turbulent fluid, but rather if one tosses both a red and a blue cork into a turbulent fluid, one can predict the distribution of the difference of the velocities of two different corks.

Perhaps the most intriguing question is "Why should economics (which arises as a result of human behavior) have anything to do with statistical physics (which studies the motion of inanimate objects, such as corks tossed into turbulent fluids)? The behavior of the cork (say 1 cm in diameter) is influenced by the collisions it suffers each picosecond from 10^{15} microscopic water molecules. However, the advances in turbulence do not arise from "summing up" on a computer these 10^{15} collisions each picosecond. Rather we "understand" what we observe in terms of a collective phenomenon which obeys not Newton's laws but rather scaling laws.

Consider a system of interest in economics, such as the stock market. Here the value of a stock index is influenced by all the traders to varying degrees, since if any group of traders decides to behave in a certain way then the value of that stock index will change. Now each stock trader is certainly not inanimate. However, each is following certain well defined rules, just as the inanimate cork follows certain rules. We don't know much about the complete microscopic connection between the macroscopic observed quantity and the individual rules followed by the stock trader. Similarly, we don't know much concerning the complete microscopic connection between the macroscopic observed quantity, the cork velocity, and the microscopic laws followed by the 10^{15} water molecules that collide with the cork each picosecond.

Our work concerns not stock market index but rather the sales of every publicly-traded firm in the USA. It is not implausible that this macroscopic variable, sales, is

influenced to varying degrees by the actions of a nonnegligible fraction of the earth's total population. Thus again we observe a macroscopic variable whose microscopic inputs are known in principle, but how these microscopic events "conspire" to set the actual sales of a given firm is probably as hopeless a problem to compute as is the motion of a cork buffeted by 10^{15} collisions each picosecond.

In summary, we are seeking methods of providing new quantitative information on a distribution function fundamental to key economic questions. We find a new set of quantitative methods for characterizing economic data – using scaling invariance theory – which holds promise for shedding new light. Indeed, the analogs of universality classes in critical phenomena seems to be within reach, since data from other economies can also be analyzed with reasonable effort. It may be that as we study economies less regulated or more regulated (by studying, e.g., data bases from East Asia, Europe, or perhaps pre-war USA) we will find as rich a phenomenology as was discovered to describe the various universality classes in critical phenomena. And it may be that theoretical models to explain our empirical findings will be forthcoming.

Acknowledgements

We wish to thank I. Grosse, H. Herzel, E. P. Kolker, M.F. Shlesinger, and E. N. Trifonov for helpful ideas at the initial stages of this work, and NIH-HGP, NSF, the Mathers Charitable Foundation, the American Heart Association, Istituto Nazionale di Fisica della Materia, and the Israel-USA Binational Science Foundation for support.

References

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

[2] C.K. Peng, J.M. Hausdorff, J.E. Mietus, S. Havlin, H.E. Stanley and A.L. Goldberger, Fractals in physiological control: from heartbeat to gait, in: Lévy Flights and Related Topics in Physics (Proc. 1994 International Conference on Lévy Flights), edited by M.F. Shlesinger, G.M. Zaslavsky and U. Frisch (Springer, Berlin, 1995), pp. 315–330;
C.K. Peng, S. Havlin, H.E. Stanley and A.L. Goldberger, Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series, in: Proc. NATO Dynamical Disease Conference, edited by L. Glass, Chaos 5 (1995) 82–87.

[3] G.M. Viswanathan, V. Afanasyev, S.V. Buldyrev, P.A. Prince and H.E. Stanley, Lévy flight search patterns in animal behavior, Nature (submitted).

[4] R.N. Mantegna and H.E. Stanley, Nature 376 (1995) 46–49.

[5] M.H.R. Stanley, S.V. Buldyrev, S. Havlin, R. Mantegna, M.A. Salinger and H.E. Stanley, Eco. Lett. 49 (1995) 453–457.

[6] M.H.R. Stanley, L.A.N. Amaral, S.V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M.A. Salinger and H.E. Stanley, Scaling behavior of firm growth, Nature (in press);
M.H.R. Stanley, L.A.N. Amaral, S.V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M.A. Salinger and H.E. Stanley, Can the science of statistical physics contribute to the science of economics?, in: Proc. International Conference on Future of Fractals, Fractals 3 (1995).

[7] S. V. Buldyrev, A. L. Goldberger, S. Havlin, C.-K. Peng and H. E. Stanley, Fractals in biology and medicine: from DNA to the Heartbeat, in: Fractals in Science, edited by A. Bunde and S. Havlin (Springer, Berlin, 1994), Chapter 2.

[8] E.W. Montroll and M.F. Shlesinger, The wonderful world of random walks, in: Nonequilibrium Phenomena II. From Stochastics to Hydrodynamics, edited by J.L. Lebowitz and E.W. Montroll (North-Holland, Amsterdam, 1984), pp. 1–121.
[9] G.H. Weiss, Random Walks (North-Holland, Amsterdam, 1994);
A.-L. Barabasi and H.E. Stanley, Fractal Concepts in Surface Growth (Cambridge University Press, Cambridge, 1995).
[10] S. Havlin, R. Selinger, M. Schwartz, H.E. Stanley and A. Bunde, Phys. Rev. Lett. 61 (1988) 1438;
S. Havlin, M. Schwartz, R. Blumberg Selinger, A. Bunde and H.E. Stanley, Phys. Rev. A 40 (1989) 1717;
R.B. Selinger, S. Havlin, F. Leyvraz, M. Schwartz and H.E. Stanley, Phys. Rev. A 40 (1989) 6755.
[11] C.-K. Peng, S. Havlin, M. Schwartz, H.E. Stanley and G.H. Weiss, Physica A 178 (1991) 401;
C.-K. Peng, S. Havlin, M. Schwartz and H.E. Stanley, Phys. Rev. A 44 (1991) 2239.
[12] M. Araujo, S. Havlin, G.H. Weiss and H.E. Stanley, Phys. Rev. A 43 (1991) 5207;
S. Havlin, S.V. Buldyrev, H.E. Stanley and G.H. Weiss, J. Phys. A 24 (1991) L925;
S. Prakash, S. Havlin, M. Schwartz and H.E. Stanley, Phys. Rev. A 46 (1992) R1724.
[13] M. Ya. Azbel, Phys. Rev. Lett. 31 (1973) 589; Phys. Rev. Lett. 75 (1995) 168–171.
[14] S.V. Buldyrev, A.L. Goldberger, S. Havlin, R.N. Mantegna, M.E. Matsa, C.-K. Peng, M. Simons and H.E. Stanley, Phys. Rev. E 51 (1995) 5084–5091.
[15] W. Li and K. Kaneko, Europhys. Lett. 17 (1992) 655.
[16] R. Voss, Phys. Rev. Lett. 68 (1992) 3805.
[17] S. Karlin and V. Brendel, Science 259 (1993) 677.
[18] S.V. Buldyrev, A.L. Goldberger, S. Havlin, C.-K. Peng, M. Simons and H.E. Stanley, Phys. Rev. E 47 (1993) 4514–4523.
[19] C.K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley and A. L. Goldberger, Phys. Rev. E 49 (1994) 1685–1689.
[20] H. Makse, S. Havlin, M. Schwartz and H.E. Stanley, Method for generating long-range correlations for large systems, Phys. Rev. E Rapid Communications (submitted); Proc. 1993 Int. Conf. Complex Systems in Computational Physics, Buenos Aires, Chaos, Solitons and Fractals 6 (1995) 295;
see also N.N. Pang, Y.-K. Yu and T. Halpin-Healy, Phys. Rev. E 52 (1995) 3224.
[21] S. Prakash, S. Havlin, M. Schwartz and H.E. Stanley, Phys. Rev. A 46 (1992) R-1724.
[22] C.K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, M. Simons and H.E. Stanley, Phys. Rev. E 47 (1993) 3730–3733.
[23] S.M. Ossadnik, S.V. Buldyrev, A. L. Goldberger, S. Havlin, R.N. Mantegna, C.-K. Peng, M. Simons and H.E. Stanley, Biophys. J. 67 (1994) 64–70;
H.E. Stanley, S.V. Buldyrev, A.L. Goldberger, S. Havlin, C.-K. Peng and M. Simons, Physica A 200 (1993) 4–24;
H.E. Stanley, S.V. Buldyrev, A.L. Goldberger, S. Havlin, S.M. Ossadnik, C.-K. Peng and M. Simons, Fractals 1 (1993) 283–301.
[24] W. Li, T.G. Marr and K. Kaneko, Physica D 75 (1994) 392–416.
[25] A. Arneodo, E. Bacry, P.V. Graves and J.F. Muzy, Phys. Rev. Lett. 74 (1995) 3293–3296.
[26] W. Li, Phys. Rev. A 43 (1991) 5240–5260;
W. Li, International Journal of Bifurcation and Chaos 2 (1992) 137–154.
[27] G.I. Bell, Computers and Chemistry 17 (1993) 185–190.
[28] G.R. Sutherland and R.I. Richards, P.N.A.S./USA 92 (1995) 3636–3641.
[29] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts and J.D. Watson, Molecular Biology of the Cell, Third Edition (Garland Publishing, New York, 1994).
[30] J. Jurka, J. Mol. Evol. 29 (1989) 496.
[31] R.H. Hwu, J.W. Roberts, E.H. Davidson and R.J. Britten, P.N.A.S./USA 83 (1986) 3875.
[32] J. Jurka, T. Walichiewicz and A. Milosevljevic, J. Mol. Evol. 35 (1992) 286.
[33] S.V. Buldyrev, A.L. Goldberger, S. Havlin, C.-K. Peng, H.E. Stanley, M.H.R.Stanley and M. Simons, Biophys. J. 65 (1993) 2673–2679.
[34] J. Des Cloizeaux, J. Physique (Paris) 41 (1980) 223;
P.G. de Gennes, Scaling Concepts in Polymer Physics (Cornell University Press, Ithaca, 1979).
[35] G.A. Churchill, Computers and Chemistry 16 (1992) 107–116.
[36] V. Afanasyev and P.A. Prince, Ornis Scandinavica 24 (1993) 243.

[37] M.F. Shlesinger and B.J. West, Phys. Rev. Lett. 67 (1991) 2106;
see also M.F. Shlesinger, Ann. NY Acad. Sci. 504 (1987) 214.
[38] A. Yu. Grosberg and A.R. Khokhlov, Statistical Physics of Macromolecules (AIP Press, New York, 1994).
[39] B.J. West and M.F. Shlesinger, Am. Scientist 78 (1990) 40.
[40] E. Canessa and A. Calmetta, Phys. Rev. E 50 (1994) R47.
[41] C.K. Peng, J. Mietus, J. Hausdorff, S. Havlin, H.E. Stanley and A.L. Goldberger, Phys. Rev. Lett. 70 (1993) 1343–1346.
[42] J.J. Collins and C.J. De Luca, Phys. Rev. Lett. 73 (1994) 764;
J.M. Hausdorff, C.-K. Peng, Z. Ladin. J.Y. Wei and A.L. Goldberger, J. Appl. Physiol. 78 (1995) 349–358.
[43] J.B. Bassingthwaighte, L.S. Liebovitch and B.J. West, Fractal Physiology (Oxford University Press, New York, 1994).
[44] B.J. West and W. Deering, Phys. Reports 246 (1994) 1.
[45] T. Vicsek, M.F. Shlesinger and M. Matsushita, eds., Fractals in Natural Sciences (World Scientific, Singapore, 1994).
[46] B.J. West, Fractal Physiology and Chaos in Medicine (World Scientific, Singapore 1990).
[47] B. Drossel, S. Clar and F. Schwabl, Phys. Rev. Lett. 69 (1992) 1629.
[48] M.F. Shlesinger, in: On Growth and Form, edited by H. E. Stanley and N. Ostrowsky (Nijhoff, Dordrecht, 1986), p. 283.
[49] M.F. Shlesinger, G.M. Zaslavsky and U. Frisch, eds., Lévy Flights and Related Topics in Physics (Springer-Verlag, Berlin, 1995).
[50] R.N. Mantegna and H.E. Stanley, Stock market dynamics and turbulence: parallels in quantitative measures of fluctuation phenomena, Nature (submitted).