



Self-affine analysis of protein energy

P.H. Figueirêdo^a, M.A. Moret^{b,c,*}, P.G. Pascutti^d, E. Nogueira Jr.^e, S. Coutinho^f

^a Departamento de Física, Universidade Federal Rural de Pernambuco, CEP 52171-900, Recife, Pernambuco, Brazil

^b Programa de Modelagem Computacional - SENAI - Cimatec, 41650-010 Salvador, Bahia, Brazil

^c Departamento de Física, Universidade Estadual de Feira de Santana, CEP 44031-460, Feira de Santana, Bahia, Brazil

^d Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

^e Instituto de Física, Universidade Federal da Bahia, CEP 40210-340, Salvador, Bahia, Brazil

^f Departamento de Física, Universidade Federal de Pernambuco, CEP 50670-901, Recife, Pernambuco, Brazil

ARTICLE INFO

Article history:

Received 14 March 2009

Received in revised form 26 February 2010

Available online 23 March 2010

Keywords:

Protein packing

Self-affinity

ABSTRACT

We study the time series of the total energy of polypeptides and proteins. These time series were generated by molecular dynamics methods and analyzed by applying detrended fluctuation analysis to estimate the long-range power-law correlation, i.e. to measure scaling exponents α . Such exponents were calculated for all systems and their values follow environment conditions, i.e., they are temperature dependent and also, in a continuum medium approach, vary according to the dielectric constants (we simulated $\epsilon = 2$ and $\epsilon = 80$). The procedure was applied to investigate polyalanines, and other realistic models of proteins (Insect Defensin A and Hemoglobin). The present findings exhibit results that are consistent with previous ones obtained by other methodologies.

© 2010 Elsevier B.V. All rights reserved.

In recent years, there has been a growing evidence that many complex physical, economical, and biological systems manifest self-affinity characterized by long-range power-law correlations. In such a context, the detrended fluctuation analysis (DFA) was recently proposed [1] to analyze long-range power-law correlations in nonstationary systems. One advantage of the DFA method is that it allows the long-range power-law correlations in signals with embedded polynomial trends that can mask the true correlations in the fluctuations of a noise signal. The DFA method has been applied to analyze the DNA and its evolution [1,2], file editions in computer diskettes [3], economics [4,5], climate temperature behavior [6], phase transition [7], astrophysics sources [8,9] and cardiac dynamics [10,11], among others.

The study of fractal characteristics of the proteins provides countless results. The fractal analysis uncovered self-similarity in many research fields such as cluster dimension of proteins [12], anomalous temperature dependence of the Raman spin-lattice relaxation rates [13], relation between the fractal dimension and the number of hydrogen bridges [14], multifractality in the energy hypersurface of the proteins [15], packing of small protein fragments [16], surface volume [17], degree of compactness of the proteins [18], measurement of the average packing density [19] as well as a hydrophobicity scale [20] among others. Furthermore, the fractal methods identify different states of the same system according to its different scaling behaviors, e.g., the fractal dimension is different for structures with (without) hydrogen bonds [14,15]. In this sense, the correct interpretation of the scaling results obtained by the fractal analysis is crucial to understand the intrinsic geometry (and sometimes dynamics) of the systems under study.

In this paper the DFA method was applied to investigate self-affinity presented in protein molecular dynamics. The long-term time-series energy of peptides and proteins were studied by the using the THOR modelling program. The THOR program

* Corresponding author at: Departamento de Física, Universidade Estadual de Feira de Santana, CEP 44031-460, Feira de Santana, Bahia, Brazil. Tel.: +55 71 82272352.

E-mail address: mamoret@gmail.com (M.A. Moret).

[21–23] was developed to be a comprehensive and a flexible tool to investigate macromolecular structures of biological interest such as proteins [15,24,25] and the interaction between biological macromolecules and cellular membranes [26].

The computational code is based on a classical force field and considers the GROMOS parameters [27]. However, other force fields can be easily implemented. Both molecular dynamics [21,23,25,26,28] and optimization methods [15,22,24,29–32] are available in this program. The choice of method depending upon the user necessity. In the THOR program, the conformational energy of the molecule is made up of a sum of bonded and nonbonded terms. In such approach, only hydrogen atoms covalently bonded to oxygen or to nitrogen are considered explicitly, whereas CH1, CH2, and CH3 groups are assumed to be an atomic unit. Thus, we analyze the changes of the following energy function:

$$E = E_H + E_\theta + E_\phi + E_\varphi + E_{LJ} + E_{el} \quad (1)$$

$$E = \frac{1}{2} \sum_k K_{b_k} (r_k + r_0)^2 + \frac{1}{2} \sum_l K_{\theta_l} (\theta_l + \theta_0)^2 + \frac{1}{2} \sum_m K_{\phi_m} (\phi_m + \phi_0)^2 \\ + \sum_n K_{\varphi_n} (1 + \cos(m\varphi_n + \varphi_0)) + \sum_{i < j} \left[\frac{C_{12}(i, j)}{r_{i,j}^{12}} - \frac{C_6(i, j)}{r_{i,j}^6} \right] + \frac{1}{4\pi \epsilon_0 \epsilon_r} \sum_{i < j} \frac{q_i q_j}{r_{i,j}} \quad (2)$$

where E_H is the Hook potential, E_θ is the angular potential, E_ϕ is the improper potential, E_φ is the dihedral angle potential, E_{LJ} is the Lennard Jones potential, and E_{el} is the Coulomb potential term (see definitions and used parameters in Refs. [21–28]).

To analyze the time-series energy (E) we present a brief introduction to the DFA method [1,2], which involves the following steps:

- (1) Consider a correlated signal-intensity series, $E(i)$ (kcal/mol), where $i = 1, \dots, N_{max}$ and N_{max} is the total number of molecular dynamics steps. The signal $E(i)$ is integrated to obtain $y(k) = \sum_{i=1}^k (E(i) - \langle E \rangle)$, where $\langle E \rangle$ stands for the average value of the energy;
- (2) The integrated signal $y(k)$ is divided into boxes of equal length n ;
- (3) For each n -size box, $y(k)$ is fitted using a polynomial function of order l , which represents the trend in the box. The y coordinate of the fitting line in each box is denoted by $y_n(k)$, since we use a polynomial fitting of order l , we denote the algorithm as DFA- l ;
- (4) The integrated signal $y(k)$ is detrended by subtracting the local trend $y_n(k)$ within each box (of length n);
- (5) For a given n -size box, the root-mean-square fluctuation, $F(n)$, for the integrated and detrended signal is given

$$F(n) = \sqrt{\frac{1}{N_{max}} \sum_{k=1}^{N_{max}} [y(k) - y_n(k)]^2}; \quad (3)$$

- (6) The above computation is repeated for a broad range of scales (n -size box) to provide a relationship between $F(n)$ and the box size n .

The scaling exponent α is defined whenever such relationship is characterized by a power-law $F(n) \propto n^\alpha$. Therefore, the scaling exponent α is a self-affine parameter expressing the long-range power-law correlation properties of the energy signal: for $\alpha < 0.5$ the correlation in the energy signal is anti-persistent, while for $\alpha > 0.5$ the correlation in the energy signal is persistent. If $\alpha = 0.5$ the energy signal is said to be uncorrelated.

Note that for anti-persistent signals, the scaling exponent obtained from the DFA method overestimates the true correlations at small scales. To avoid this problem, one needs first to integrate the original anti-correlated signal and then apply the DFA method. The correct scaling exponent can thus be obtained from the relation between n and $F(n)/n$ [instead of $F(n)$]. Therefore when the signal was anti-persistent we estimated α using this methodology [33]. The DFA analysis was carried out over the total energy series of molecules modelled during 1 ns, and all the systems show a power-law with one characteristic long-range correlation scaling exponent α .

Fig. 1 depicts a characteristic behavior of scaling exponent α of the analyzed time series.

Recently some fractal behaviors provide insights about protein structure, e.g., both surface volume [17] and mass-size [18] fractal analysis suggest that protein packing behaves as random spheres in percolation threshold [34].

Some macromolecules were used to investigate the scaling exponent α . Fig. 2 depicts the scaling exponent α for the small 6-alanine peptide as function of temperature for two different dielectric constants. We observe that the scaling exponent α has a persistent behavior. On the other hand, this plot also shows that the scaling exponent α for the time-series energy of this small peptide is quite independent of the Coulomb (electric) energy (last term of (2)), because the behavior of α is the same in $\epsilon = 2$ and $\epsilon = 80$ for low and high temperatures. In other words, this small theoretical molecule does not form hydrogen bonds, as observed in Ref. [24] using other methodology of conformational search.

Polyalanines are theoretical molecules largely used to modelling α -helices [15,22,24,32]. Helices are the most prevalent secondary structural motif observed in proteins with known structure. Several recent theoretical studies of the helix-coil transitions have shown the relation between the unfolded state, in random coil conformations, and the helical conformational states [24].

The α -helix is the classic element of protein structure. All the hydrogen bonds of the α -helix backbone are aligned along the helical axis with the same orientation. Because a peptide bond has a dipole moment arising from the different polarity of

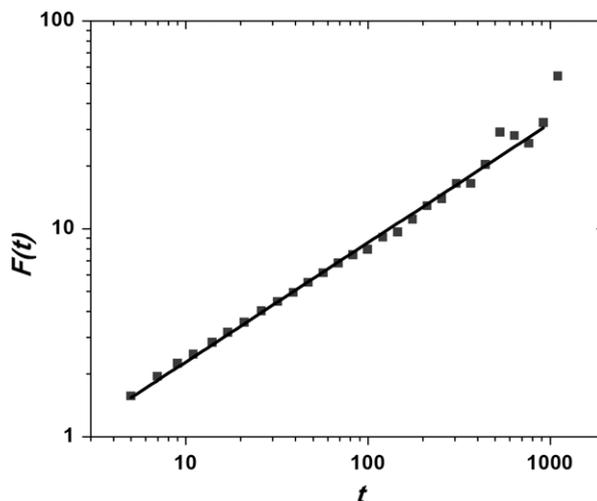


Fig. 1. A typical DFA result to protein dynamics analysis ($T = 250$ K and $\epsilon = 2$). This 20-alanine peptide obeys the scaling exponent $\alpha = 0.57 \pm 0.01$, with the correlation coefficient of Pearson $R = 0.98$.

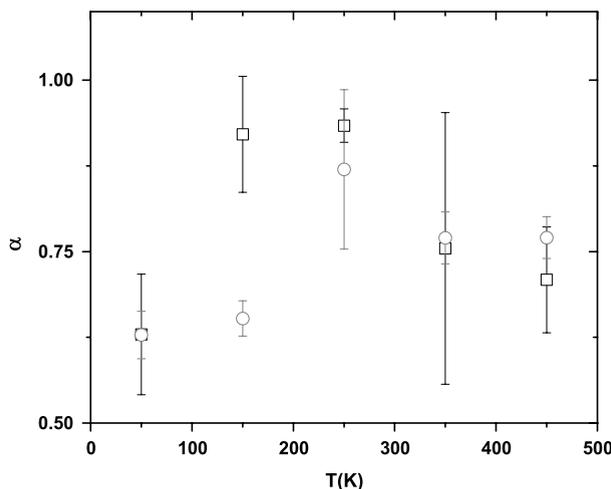


Fig. 2. The scaling exponent α as a function of the temperature for the 6-alanine peptide. Open black squares label the results for low dielectric constant ($\epsilon = 2$), while gray open circles denote the results for high dielectric constant ($\epsilon = 80$).

the NH and CO groups, these dipole moments are also aligned along the helical axis. The overall effect is a significant macro-dipole that has the positive pole at the amino end and the negative pole at the carboxyl end of the α -helix. The overall energy that stabilizes the α -helix came from the attractive contributions due to hydrogen bonds and/or by charge-helix dipole interactions [35] as well as from the van der Waals contributions. The α -helix formation is a cooperative process where the electrostatic energy has an important role in stabilizing such type of structure. A critical number of amino acids, however, are necessary to stabilize this α -helix structure, and an upper limit may also be imposed by the entropy effect. Isolated α -helix structures, in fact, would have to be longer than 13 residues to be stabilized by the attractive interactions [24].

Different amino acids have been found to present weak though definite preference in favor or against being in α -helix structure, and the intrinsic helical propensity of some amino acids has been demonstrated to be position dependent [36]. In this sense, Ala, Glu, Leu, and Met are considered to be good α -helix promoters whereas Pro, Gly, Tyr, and Ser are considered to be poor ones [37]. Such preferences were the main considerations in all early attempts to predict secondary structures from amino acid sequences, but they were not strong enough to obtain accurate predictions.

Fig. 3 presents a typical pattern of self-affine behavior for polyalanines with 13 or more residues. The plot shows the 18-alanine peptide as function of the temperature for two values of the dielectric constant. We recall that this polypeptide tends to stabilize its structure by forming hydrogen bonds, which are arduous to form when the dielectric constant is $\epsilon = 80$. Therefore, hydrogen bonds interfere in the scaling exponent α , the energy time series becoming more persistent than that when $\epsilon > 2$. Notice that for the dielectric constant $\epsilon = 2$, 13-alanine has α -helix as a the more stable structure [24].

Fig. 4 depicts the 14-alanine and 18-alanine peptides as a function of the temperature. We note that that the persistent behavior follows the formation of the hydrogen bonds. If these bonds increases ($T < 300$ K) α decreases and vice-versa.

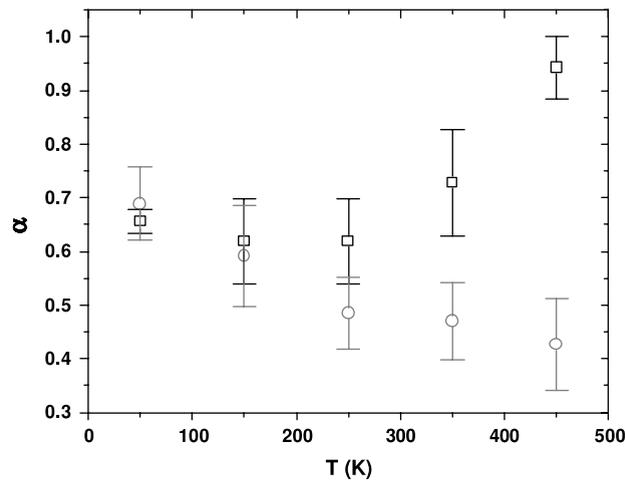


Fig. 3. The scaling exponent α as a function of the temperature for the 18-alanine peptide. Open black squares label the results for low dielectric constant ($\epsilon = 2$), while open gray circles denote the results for high dielectric constant ($\epsilon = 80$).

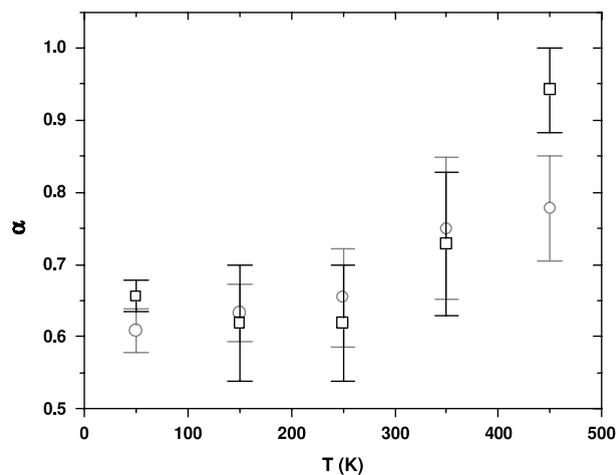


Fig. 4. The scaling exponent α for two polyanalines, 14-alanine peptide (gray circle) and 18-alanine peptide (black square). Using dielectric constant $\epsilon = 2$ the scaling exponent increases for temperatures greater than 300K.

To make comparisons to more realistic models of biological molecules we consider the Insect Defensin and the human Hemoglobin. Insect Defensin A, is a basic 4 kDa protein. This small protein presents one α -helix and two β -strands, stabilized with three disulfide bridges. These disulfide bridges maintain this molecule always close to the native structure. Also, we study the self-affinity present in the time series of the human Hemoglobin.

Fig. 5 depicts the main characteristic of all systems discussed here, using DFA analysis. The plot shows that the scaling exponent α tends to increase with the temperature indicating that, for continuous media with dielectric constant $\epsilon = 2$, the molecular system becomes more persistent when temperature increases.

We recall that values found for the scaling exponent α indicate the presence of statistical fluctuations that depend on local characteristics throughout the series and therefore suggests a multifractal analysis of energy profiles. In fact, a methodology based on Eq. (3), called multifractal detrended fluctuation analysis (MF-DFA) [38] confirms the existence of complete spectra of different exponents $\alpha(q)$ (data not shown). In particular, the analysis presented in this paper refers to the case $q = 2$, i.e., we present the main scale possible.

In summary, some aspects of the rugosity of the time series of the total energy of proteins and polypeptides were investigated by means of the detrended fluctuation analysis (DFA). The findings suggest that the correlated self-affine parameter (scaling exponent α) is temperature dependent and it vary with the dielectric constant. This result is in agreement to previous one obtained by Ref. [14] using a lattice model. The results for the lattice model indicate that the system is fractal, which fractal dimension depends on the number of H-bonds [14]. Using a more precise model Ref. [15] has observed that, when the system analyzed has hydrogen bonds, the multifractal $f(\alpha)$ spectrum shape is different in comparison to those systems that do not produce hydrogen bonds. These results show that the fractal dimension tends to increase with the temperature of these molecules.

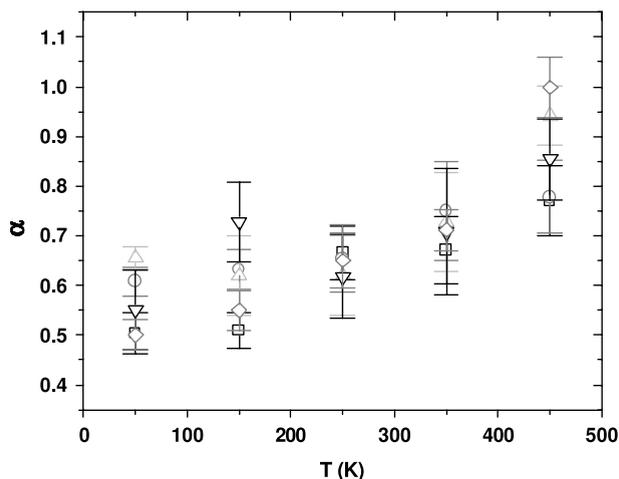


Fig. 5. The scaling exponent α as a function of the temperature for: 10-alanine (black square), 14-alanine (gray circle), 18-alanine peptide (light gray up triangle), Insect Defensin A (gray diamond) and hemoglobin (black down triangle).

From all the simulated biological systems the scaling exponent α increases for growing values of the temperature. This means that the energy hypersurface rugosity increases, due to the growth of the normal modes of vibration that are quite sensitive to temperature variations. We also found that α tends to increase when the dielectric constant decreases. Finally, for high temperatures the unrealistic water medium with dielectric constant $\epsilon = 80$ shows an anti-persistent behavior.

Acknowledgements

This work received financial support from CNPq and CAPES (Brazilian federal grant agencies) and from FACEPE (Pernambuco state grant agency under the grant PRONEX EDT 0012-05.03/04).

References

- [1] C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, *Phys. Rev. E* 49 (1994) 1685.
- [2] C.-K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, F. Sciortino, M. Simons, H.E. Stanley, *Nature (London)* 356 (1992) 168.
- [3] G.F. Zebende, P.M.C. de Oliveira, T.J.P. Penna, *Phys. Rev. E* 57 (1998) 3311.
- [4] R.N. Mantegna, H.E. Stantey, *Nature (London)* 367 (1995) 46.
- [5] A.N.S. Filho, G.F. Zebende, M.A. Moret, *Internat. J. Modern Phys. C* 19 (2008) 665.
- [6] P. Talkner, R.O. Weber, *Phys. Rev. E* 62 (2000) 150.
- [7] G.F. Zebende, M.V.S. da Silva, A.C.P. Rosa, A.S. Alves, J.C.O. de Jesus, M.A. Moret, *Physica A* 342 (2004) 322.
- [8] M.A. Moret, G.F. Zebende, E. Nogueira, M.G. Pereira, *Phys. Rev. E* 68 (2003) 041104.
- [9] G.F. Zebende, M.G. Pereira, E. Nogueira, M.A. Moret, *Physica A* 349 (2005) 452.
- [10] P.Ch. Ivanov, M.G. Rosenblum, C.-K. Peng, J.E. Mietus, S. Havlin, H.E. Stanley, A.L. Goldberger, *Nature (London)* 383 (1996) 323.
- [11] P.Ch. Ivanov, L.A. Nunes Amaral, A.L. Goldberger, S. Havlin, M.G. Rosenblum, Z.R. Struzik, H.E. Stanley, *Nature (London)* 399 (1999) 461.
- [12] T. Jossang, J. Feder, E. Resenqvist, *J. Chem. Phys.* 120 (1984) 1.
- [13] J.T. Colvin, H.J. Stapleton, *J. Chem. Phys.* 82 (1985) 4699.
- [14] J.S. Helman, A. Coniglio, C. Tsallis, *Phys. Rev. Lett.* 53 (1984) 1195.
- [15] M.A. Moret, P.G. Pascutti, K.C. Mundim, P.M. Bisch, E. Nogueira, *Phys. Rev. E* 63 (2001) 020901.
- [16] T.P. Creamer, R. Srinivasan, G.D. Rose, *Biochemistry* 34 (1995) 16245.
- [17] J. Liang, K. Dill, *Biophys. J.* 81 (2001) 751.
- [18] M.A. Moret, J.G.V. Miranda, E. Nogueira, M.C. Santana, G.F. Zebende, *Phys. Rev. E* 71 (2005) 012901.
- [19] M.A. Moret, M.C. Santana, E. Nogueira, G.F. Zebende, *Physica A* 361 (2006) 250.
- [20] M.A. Moret, G.F. Zebende, *Phys. Rev. E* 75 (2007) 011920.
- [21] E.P.G. Arêas, P.G. Pascutti, S. Schreier, K.C. Mundim, P.M. Bisch, *J. Phys. Chem.* 99 (1995) 14882.
- [22] M.A. Moret, P.G. Pascutti, P.M. Bisch, K.C. Mundim, *J. Comput. Chem.* 19 (1998) 647.
- [23] P.G. Pascutti, L.J. El-Jaik, P.M. Bisch, K.C. Mundim, A.S. Ito, *Eur. Biophys. J.* 28 (1999) 499.
- [24] M.A. Moret, P.M. Bisch, K.C. Mundim, P.G. Pascutti, *Biophys. J.* 82 (2002) 1123.
- [25] F.L. Sirota, P.G. Pascutti, C. Anteneodo, *Biophys. J.* 82 (2002) 1207.
- [26] P.G. Pascutti, K.C. Mundim, A.S. Ito, P.M. Bisch, *J. Comput. Chem.* 20 (1999) 971.
- [27] W.V. van Gunsteren, H.J.C. Berendsen, *Groningen Molecular Simulation (GROMOS) Library Manual*, Biomos, Groningen, The Netherlands, 1987.
- [28] P.R. Batista, A.W.S. DaSilva, E.H.A. Barbosa, P.G. Pascutti, *Cell Biochem. Biophys.* 44 (2006) 395.
- [29] M.A. Moret, P.M. Bisch, F.M.C. Vieira, *Phys. Rev. E* 57 (1998) R2535.
- [30] M.A. Moret, P.M. Bisch, E. Nogueira, P.G. Pascutti, *Physica A* 353 (2005) 353.
- [31] M.A. Moret, P.G. Pascutti, P.M. Bisch, M.S.P. Mundim, K.C. Mundim, *Physica A* 363 (2006) 260.
- [32] F.P. Agostini, D.D.O. Soares-Pinto, M.A. Moret, C. Osthoff, P.G. Pascutti, *J. Comput. Chem.* 27 (2006) 1142.
- [33] Z. Chen, P.C. Ivanov, K. Hu, H.E. Stanley, *Phys. Rev. E* 65 (2002) 041107.
- [34] D. Stauffer, *Phys. Rep.* 54 (1979) 1.
- [35] K.R. Shoemaker, P.S. Kim, D.N. Brems, S. Marqusee, E.J. York, I.M. Chaiken, J.M. Stewart, R.L. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.* 82 (1985) 2349.
- [36] M. Petukhov, K. Uegaki, N. Yumoto, S. Yoshikawa, L. Serrano, *Protein Sci.* 8 (1999) 2144.
- [37] C. Branden, J. Tooze, *Introduction to Protein Structure*, Garland Publishing, New York, 1991.
- [38] J.W. Kantelhardt, S.A. Zschiegner, E. Koncsieny-Bunde, S. Havlin, A. Bunde, H.E. Stanley, *Physica A* 316 (2002) 87.