Physica A 415 (2014) 440–448

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Fractals and self-organized criticality in proteins

J.C. Phillips *

Department of Physics and Astronomy, Rutgers University, Piscataway, NJ, 08854, United States

HIGHLIGHTS

- Fractals are a powerful tool for holistic analysis of proteins.
- The globular structure of proteins is fixed by their surfaces and by the folded protein chain.
- The differential geometry of globular proteins is fixed by hydropathic forces.
- A new Brazilian scale accurately quantifies these forces using fractals.
- Protein evolution reflects functions, identified with no adjustable parameters

ARTICLE INFO

Article history:
Received 11 March 2014
Received in revised form 10 July 2014
Available online 19 August 2014

Keywords:
Hydrophobic
Hydrophilic
Surface area
Evolution
Sequence
Scaling

ABSTRACT

Self-organized criticality, a powerful concept, originated in 1987 as an extension of fractal geometries to thermodynamic systems in the vicinities of instabilities. The value of fractal methods can be greatly enhanced in realistic models that exploit accurate fractal values derived from homogeneous (possibly curated) Big Data. We illustrate this point by discussing the derivation of fractal exponents describing protein–water interactions, and their application to protein roughness, protein binding and potentially protein engineering. The examples studied are evolution of lysozyme *c* and acylphosphatase, and mutational effects on their aggregation.

1. Introduction

Many physical systems exhibit power-law distributions over limited ranges (hence the enduring popularity of log–log plots), and power-law distributions are the characteristic feature of the modern theory of phase transitions near a critical point [1]. Self-organized criticality (SOC) is a methodology that attempts to explain why so many complex systems exhibit power-law distributions and appear to be “accidentally” located near critical points. It is argued that the critical points are dynamical fixed points (“tipping points”) towards which the system evolves without tuning external parameters [2]. The critical points are extrema in some property (or properties) with respect to which the system has been nearly optimized, especially with respect to long-range, highly cooperative interactions. SOC explains simply and quite generally the power-law distributions that are observed in many complex self-organized networks over 6 decades. It formalizes common experience as described by the law of diminishing returns, and rationalizes universality as a result of network density percolation [3].

The scope of this “unfamiliar” titled subject is indicated by its numbers relative to the overall field of fractals (40,000 papers). Self-organized criticality has 3600 papers, and is especially useful in describing networks (500 papers). Proteins

* Tel.: +1 9082738218.
E-mail address: jcphillips8@comcast.net.

http://dx.doi.org/10.1016/j.physa.2014.08.034

and fractals are discussed in 1500 papers. Another large field is neural networks (65,000 papers), which are appealing as models for the construction and/or programming of computers [4]. An infant research field (30 papers) combines SOC and neural networks [5].

Here our attention is focused on exploiting the recent discovery of fractal SOC in the differential geometry of solvent-exposed surface areas of amino acids (aa) in protein sequences [6]. These areas are computed from the very extensive structures contained in the Protein Data Base, and utilized >5000 high-resolution protein segments. The areas are computed from aa spheres with Van der Waal radii, truncated in overlapping regions according to the 17th century Voronoi procedure. Small solvent-exposed areas are associated with hydrophobic aa, reflecting their tendencies to be in the globular interior, while the much larger areas associated with hydrophilic aa, reflect their tendencies to be on the folded globular protein surface.

The new feature of Ref. [6] that has profound implications for protein science, ranging from evolution to engineering, is the discovery of physically significant protein fractals. Let $\phi(\text{aa})$ represent the solvent-accessible (to a 2 Å water sphere) surface area of aa in a given structure of an entire protein, represented by a string of usually hundreds of letters, each letter representing one aa of the protein, drawn from a menu of 20 letters (aa). One then partitions the entire protein into a set of segments of length $2N + 1$, and plots $\phi(\text{aa}, N)$ against N. This function is averaged over >5000 high-resolution protein segments for each symmetrically centered aa. $\text{Log}\phi(\text{aa}, N)$ against $\log N$ shows linearity for all 20 aa for $4 \leq N \leq 17$, that is, over 0.6+ decades [6].

By itself this fractal linearity for a single aa is not impressive, and it could be accidental. However, the same linearity is obtained, over the same range, for all 20 aa (but with different fractal slopes, denoted by $\psi(\text{aa})$). The center of this common range is $N^* = 10$, and $2N^* + 1 = 21$ is a typical membrane thickness. This suggests that modular protein building blocks derived from $4 \leq N \leq 17$ have evolved as a result of protein interactions with cellular boundaries. This means that the 20 fractals of Ref. [6] constitute a robust set of parameters compressed by evolution. Of course, each $\psi(\text{aa})$ is a fractal, and taken serially over the $M = 20$ dimensional space defined by $M = 20$ aa, they could correspond to a single fractal spanning 0.6 M $\sim$ 10+ decades. This crude estimate is not a record; data for the same material (amorphous Se) fit stretched exponential relaxation with a fractal exponent of 3/7 over 12 decades [7]. The 3/7 fractal was predicted in 1994 from a survey of curated glass relaxation data, and confirmed in 2013 by Corning scientists on modern, highly homogeneous glass plates [8]. One can make a more accurate estimate of a lower hierarchical ordering bound for the equivalent single protein fractal range from information theory, which suggests that the number of equivalent decades spanned is at least $\log(17/4)(M \ln M - M)$, with $M = 20$, or $\sim$25 decades, surely a record.

The physical significance of protein dimensionless fractals is brought out by comparison with hydrophobicities determined by enthalpy differences of synthetic amino acid chains from water to air, the standard first-order scale labeled KD (15K citations) [9]. This scale corresponds to a first-order (water–air) phase transition, and it could describe protein functions associated with drastic changes in protein structure. The correlation coefficient of the two scales is 0.85, which is reasonable considering that the fractal scale describes second-order (small conformational or collective) changes, while the first-order scale approximates protein unfolding [10]. The biophysical significance of the fractals in the MZ hydropathicity scale [6,10] is indisputable.

Using fractals, modern statistical mechanics quantifies many classical properties – such as critical opalescence (discovered, 1869, explained with mean fields by M. Smoluchowski, 1908, and A. Einstein, 1910, see Internet for videos) – with high accuracy. While these connections are familiar to mathematicians and to most physical scientists, they appear to be unknown to most biologists, and are never used by them. Here we will show through examples the benefits of fractals and SOC in protein science. Critical opalescence couples long wave length light waves to long wave length density fluctuations. Similarly, SOC couples long wave length ice-like film waves to long wave length (conformational) solvent-exposed protein area fluctuations.

2. Fractal power—lysozyme examples

Detailed applications of the fractal MZ scale [6] to specific proteins often yield exponentially dramatic results, far beyond the reach of polynomial molecular dynamics simulations, no matter how large the supercomputer – or even computer "cloud" – used. Moreover, these calculations are readily carried out on EXCEL. Thus one need not be "chained" to any specific supercomputer program or even family of proteins. The fractal hydropathic scale [10] is universal, and the thermodynamic significance of results obtained using first- and critical second-order mechanics is easily seen by comparing results obtained from both the first- and second-order hydropathicity scales [8,9] comparably listed in Table 2 of Ref. [10].

The first example discussed here is lysozyme *c* (aka Hen Egg White, HEW), a comparatively small 148 aa protein, inherited mutations of which are associated with amyloidosis—Alzheimer's disease. HEW is very well studied—the PDB contains more than 200 human and 400 chicken HEW structures. HEW is also present in many other species, not only in the 400 million year old chicken sequence, but also in most other vertebrates, almost unchanged in its backbone structure. The backbone structure is exceptionally stable, with human and chicken C_α positions superposable to 1.5 Å, while the aa sequence mutates from chicken to human with 60% aa conservation [10], well above the 40% necessary for Euclidean fold conservation [11].

From chicken to human, HEW exhibits opposing enzymatic and antibiotic trends [10]. Here we focus on quantifying the simplest hydropathic features of HEW. It consists of three parts, which following Uniprot P61626 we number from 1 to 148 (more often the first 18 aa are omitted). The three parts consist primarily [12] of rigid α helices (1–56 and 104–148) and