Cell. Mol. Life Sci. (2011) 68:2711–2737
DOI 10.1007/s00018-011-0722-6

Cellular and Molecular Life Sciences

REVIEW

Fractal symmetry of protein interior: what have we learned?

Anirban Banerji · Indira Ghosh

Received: 6 January 2011 / Revised: 21 April 2011 / Accepted: 3 May 2011 / Published online: 26 May 2011

Abstract The application of fractal dimension-based constructs to probe the protein interior dates back to the development of the concept of fractal dimension itself. Numerous approaches have been tried and tested over a course of (almost) 30 years with the aim of elucidating the various facets of symmetry of self-similarity prevalent in the protein interior. In the last 5 years especially, there has been a startling upsurge of research that innovatively stretches the limits of fractal-based studies to present an array of unexpected results on the biophysical properties of protein interior. In this article, we introduce readers to the fundamentals of fractals, reviewing the commonality (and the lack of it) between these approaches before exploring the patterns in the results that they produced. Clustering the approaches in major schools of protein self-similarity studies, we describe the evolution of fractal dimension-based methodologies. The genealogy of approaches (and results) presented here portrays a clear picture of the contemporary state of fractal-based studies in the context of the protein interior. To underline the utility of fractal dimension-based measures further, we have performed a correlation dimension analysis on all of the available non-redundant protein structures, both at the level of an individual protein and at the level of structural domains. In this investigation, we were able to separately quantify the self-similar symmetries in spatial correlation patterns amongst peptide–dipole units, charged amino acids, residues with the π-electron cloud and hydrophobic amino acids. The results revealed that electrostatic environments in the interiors of proteins belonging to 'α/α toroid' (all-α class) and 'PLP-dependent transferase-like' domains (α/β class) are highly conducive. In contrast, the interiors of 'zinc finger design' ('designed proteins') and 'knottins' ('small proteins') were identified as folds with the least conducive electrostatic environments. The fold 'conotoxins' (peptides) could be unambiguously identified as one type with the least stability. The same analyses revealed that peptide–dipoles in the α/β class of proteins, in general, are more correlated to each other than are the peptide–dipoles in proteins belonging to the all-α class. Highly favorable electrostatic milieu in the interiors of TIM-barrel, α/β-hydrolase structures could explain their remarkably conserved (evolutionary) stability from a new light. Finally, we point out certain inherent limitations of fractal constructs before attempting to identify the areas and problems where the implementation of fractal dimension-based constructs can be of paramount help to unearth latent information on protein structural properties.

Electronic supplementary material The online version of this article (doi:10.1007/s00018-011-0722-6) contains supplementary material, which is available to authorized users.

A. Banerji
Bioinformatics Centre, University of Pune,
Pune 411007, Maharashtra, India
e-mail: anirbanab@gmail.com

Present Address:
A. Banerji · I. Ghosh
School of Information Technology,
Jawaharlal Nehru University, New Delhi, India

I. Ghosh (✉)
School of Computational & Integrative Sciences,
Jawaharlal Nehru University, New Delhi, India
e-mail: indirag@mail.jnu.ac.in; ighdna@yahoo.com

Introduction (why a fractal description of the protein interior is appropriate)

Proteins are enigmatic entities. They have been described as 'complex mesoscopic systems' [1]. Apart from being characterized by an enormous number of degrees of

freedom, proteins have multidimensional potential energy surfaces. It is a known fact [2, 3] that while a protein needs to adhere to its specific native fold to ensure its (thermal) stability, this native fold should allow certain parts of the protein to undergo large-amplitude movements. These movements (studied under the broad umbrella of fluctuation studies) assume a critical character in ensuring the desired functioning of a protein. It is equally intriguing to notice that although the structure of a protein can be accurately studied as that of a composite macromolecule, it has many a properties in common with that of macroscopic microparticles [4]. From somewhat a different perspective, crystallography studies [5] have reported on the mesoscopic nature of protein structures. The standard 'compact object description' of proteins (characterized by small-amplitude vibrations and by a low-frequency Debye density of states) cannot account for their 'non-idealistic' behavior [2, 3]. Indeed, the non-constancy of distance between any two atoms ($|r_I - r_j| \neq$ constant) in any biologically functional protein can easily be verified with the simplest of computer programs. In addition, recent [6] and previous [7] characterizations of inhomogeneous distributions of mass and hydrophobicity in the protein interior merely serve to complicate an effort to construct a straightforward and linear scheme for the description of the protein interior.

Thus, it is unrealistic to expect that constructs that are otherwise adequate to describe the complexity of simple structures (spheres, cubes, other regular structures of idealistic shape and characteristics) will be facilitate attempts to describe proteins. Clearly, to achieve this goal one needs to think beyond the tool-set of simple geometry and the physics of idealistic systems. To be categorical, one needs to adopt an approach that has the capabilities to describe the inherently inhomogeneous and nonlinear behaviors of protein structural parameters as an innate capacity of proteins, rather than treating them as deviations from (illusory) ideality. In this context, one notes that quantification of the self-similarity prevalent in protein properties can serve as a potent tool to achieve the aforementioned challenging task. Self-similarity has a unique advantage over many other possible constructs because an objective evaluation of self-similarity will enable deciphering of the hidden symmetry that connects global patterns of macroscopic properties in proteins (such as hydrophobicity distribution, polarizability distribution, etc.) with the local (atomic) interactions that produce them.

During attempts to describe many natural phenomena, researchers found that a measure denoted as fractal dimension (FD) could reliably quantify self-similarity. Hence, resorting to FD to quantify the self-similarity prevailing in the distributions of protein biophysical and biochemical properties does not appear to be unreasonable. As a result, investigations of protein structures using FD-based measures have generated enormous interest. The aim of this review is multifold: (1) to collate the various FD-based approaches that have been employed to study the protein interior over the years into distinct clusters; (2) to present an account of cases where FD-based methodologies have successfully contributed to research of protein interior. In this context, a thorough documentation of accurate predictions made from the spectrum of FD-based studies will be presented. However, fractals are no panacea, and they cannot suggest magic solutions to all the problems. Consequently, a third aim of our review is to examine the intrinsic limitations of FD-based measures. Finally, with a balanced assessment of the entire framework, we will attempt to identify some of the outstanding questions for which the application of FD-based investigations may help in deciphering deep and unexpected facets of protein structural organization.

Definition of FD

The concept of FD is based primarily upon the quantitative expression of the symmetry of self-similarity. While the Euclidean perception of symmetry revolves (essentially) around translation, rotation and reflection, self-similarity reveals itself through the symmetry of scale-invariance. This invariance implies that when any part of a self-similar object is magnified, it is seen to bear an exact resemblance to its whole. For example, a magnified view of one part of the coastline may not be a precise reproduction of the full picture, but it will definitely have the same qualitative appearance. Hence, it can be asserted that a coastline demonstrates statistical self-similarity. By definition, any structure possessing a self-similar motif that is invariant under a transformation of scale and which cannot be described by topological dimensions may be represented by a FD. For regular structures, self-similarity (in general) is geometric (the so-called 'geometric self-similarity'); for random or irregular objects, however, one observes self-similarity to be statistical in nature. Statistical self-similarity implies that the statistics of the spatial correlations in the patterns associated with the description of the system are similar across a range of length scales. See Fig. 1 for a description of the differences between geometric and statistical self-similarity. Here we note that the average end-to-end length of an unbranched polymer chain demonstrates statistical self-similarity [8–10] (this point will be returned to later in this review).

The dimension of an object measures its complexity. For example, in Euclidean space, the dimension of set measures the amount of information that is necessary and sufficient to identify any point of that set. Traditionally,

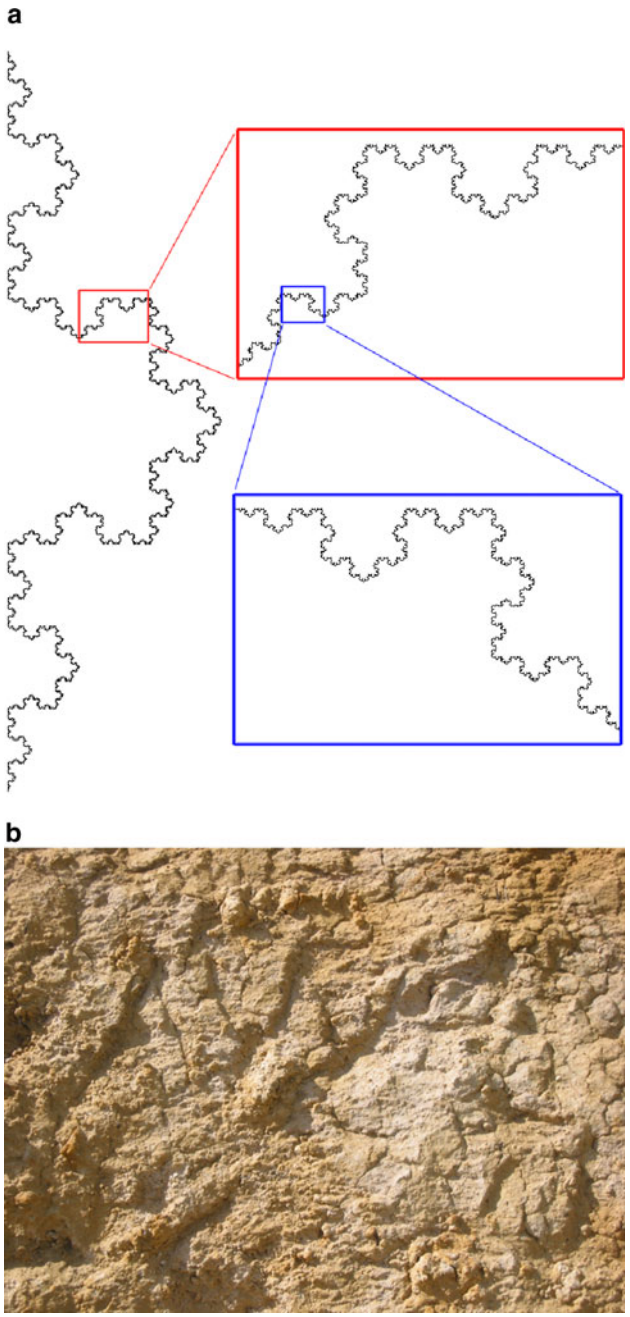

Fig. 1 **a** Example of exact (geometrical) self-similarity: the 'Koch-curve'. **b** Example of statistical self-similarity: roughness profile of the Earth's surface. Both diagrams have been reproduced courtesy of Professor Paul Bourke [153]

points have zero dimension as there is nothing to specify, straight lines and curves have one dimension because they can be parameterized by a single real coordinate and surfaces have two dimensions since two coordinates can describe every point on a surface. However, this intuitive concept of dimension cannot adequately characterize many 'real-world' objects [11] that can be characterized by the fractal dimensions. Many researchers tend to refer to the same as the 'Hausdorff dimension' [12] because Mandelbrot's definition provides a simplification of a more formal concept proposed by Hausdorff in 1919 [13]. For a formal introduction to fractals, the reader is referred to Falconer [14] and Meakin [15].

Schools of protein interior fractal studies

Application of FD-based methodology to investigate proteins dates back to (almost) the beginning of the development of modern fractal theory. This historical perspective helps us to understand why two of the most fundamental forms of protein information, namely, the primary structure and three-dimensional structure of proteins, respectively [16–18], were chosen for study in preliminary investigations. The school of primary structure studies owes its origin to polymer research and focuses primarily upon examining fractal properties within backbone connectivity. The other school of studies largely ignored main-chain information and concentrated on the distribution of residues and atoms within protein space. These schools provided a general framework by which to quantify interior fractals. Various sub-approaches can be observed within both these broad-based approaches that differ from each other (mainly) in terms of their implementation strategy. There are also two other basic approaches—slightly more abstract—that have been used to probe the protein interior. One of these revolves around the application of correlation dimension-based constructs to explore interior structural invariants, and the other attempts to relate renormalization group theory to the context of protein structure analysis. It is necessary to (briefly) survey these lines of thoughts when comparing and contrasting protein structure. Such a comparative analysis helps us to assess the possibilities of implementing the approaches in particular cases while investigating the complex (nonlinear) dependencies of the protein interior. Figure 2 presents a schematic view of the families of FD-based approaches. In this review, we will not survey the various methodologies and results associated to multifractal description of protein properties.

Approaches with backbone connectivity

Based on known data and theories (see section Definition of FD dimension), in 1982 Havlin and Ben-Avraham [8–10] studied the fractal properties of polymer chains. Denoting the average end-to-end length of an unbranched polymer by L, the number of monomer segments (of length ε) by N and fractal dimension as FD, they found the following scaling relationship: $N = \left(\frac{L}{s}\right)^{\frac{1}{\upsilon_F}} = \left(\frac{L}{s}\right)^{FD}$, in which the exponent turns out to be the inverse Flory constant υ_F in polymer theory.

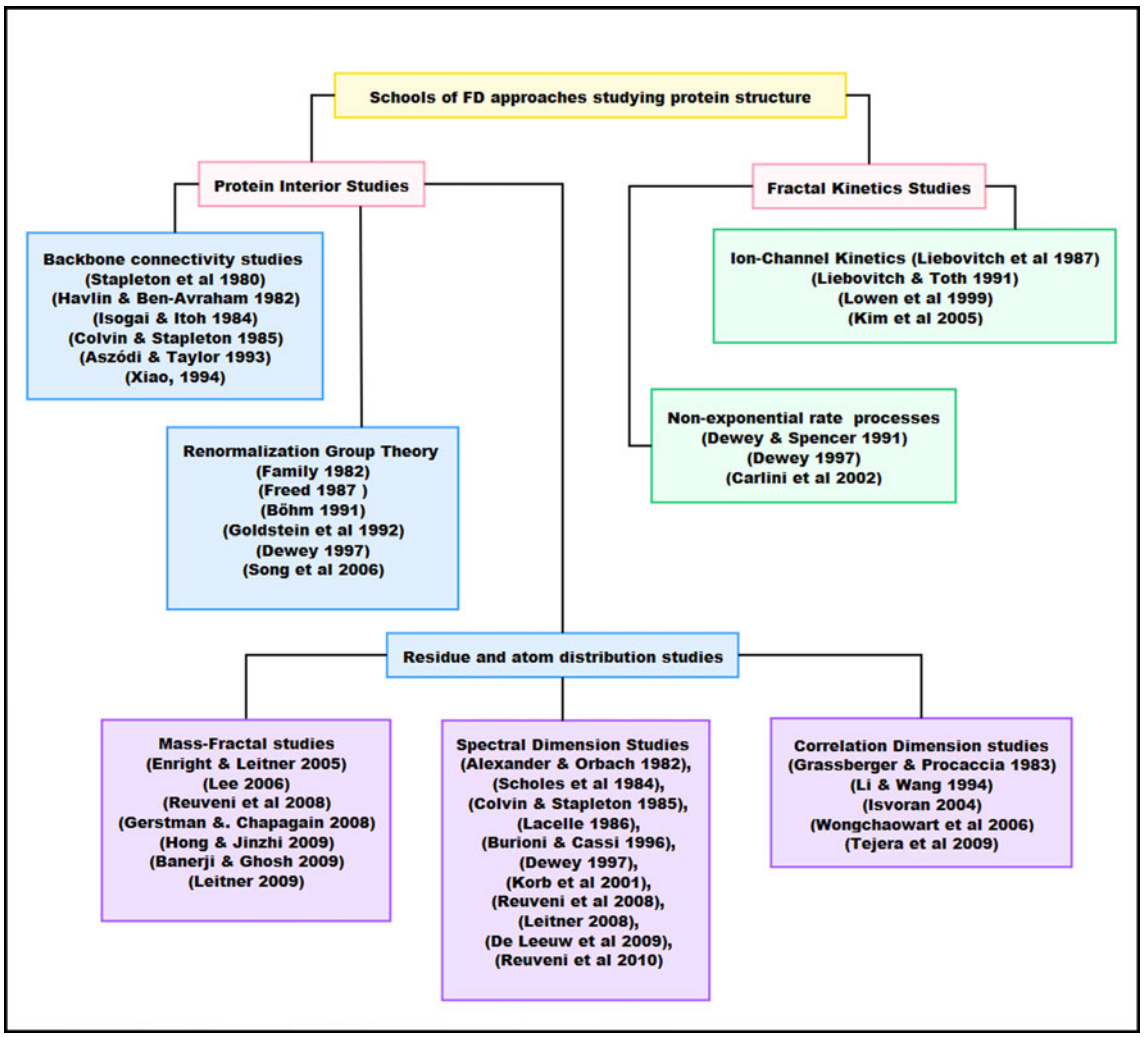

Fig. 2 Genealogy of the family of fractal-based approaches to study protein interior. *FD* Fractal dimension

(Theoretical considerations provide limits $1 < \mathrm{FD} < 2$, which correspond to a linear structure given by $L = \varepsilon N$ and to a structure represented by unrestricted random walk, when FD approaches the topological dimension magnitude 2, where $L = \varepsilon N^{\frac{1}{2}}$ [19]).

Two similar approaches emerged from this basic platform. Both considered protein length to be the most important variable, but they differed on the exact definition of protein length. The first method attempted to find the scaling exponent of the contour length, which is proportional to the number of monomers with respect to the end-to-end length. The second method attempted to quantify the length of a line segment along the chain, measured with a scale of the number of monomers in the segment. Because of its algorithmic strategy, the second method is alternatively denoted as 'the internal line segment method'.

In the first method [8, 9, 17, 18], protein length is defined as:

$$\langle Z(n)^2 \rangle = \frac{1}{N-n+1} \sum_{i=1}^{N-n+1} Z_{i,i+n}^2$$

where $Z_{i,i+n}^2$ is the square of distance between terminal residues i and $i + \mathrm{n}$ of the chain with N amino acids and n intervals. The FD in this case is calculated by performing a linear regression in the scale-range appropriate to the logarithmic form of:

$$\left[\langle Z(n)^2 \rangle^{\frac{1}{2}}\right]^{\mathrm{FD}} = \mathrm{Const}.n$$

(The exact scheme of implementation of this strategy is a bit more complex; the interested reader can consult the work of Havlin and Ben-Avraham [8, 9] for details of the algorithm and to observe the importance of FD when studying polymers).

The second method, although based on regression for the calculation of FD magnitude, differs from the first one in its definition of protein chain length. This approach [16, 20] defined protein length as:

$$L(n) = L(m) + \frac{N - n\mu - 1}{n\mu} L(m)$$

where N is the number of amino acids, n is the total length interval, $L(m)$ corresponds to the length of the chain for the first m-integer segments and $\mu = \left[\text{int}\left(\frac{N}{n}\right) - 1\right]$. The second term quantifies the remaining $(n - m)$ length of the segment. Qualitatively, this equation implies that with this scheme the emphasis rests primarily on quantifying the scaling behavior of the segmental length of the backbone, which is defined as the sum of stepwise connections of straight lines, measured for different intervals of monomers.

FD can then be obtained by analyzing $L(n) = \text{Const.} n^{1-\text{FD}}$ (only) in the scale-range where ln $[L(n)]$ and ln (n) are linearly related. Such linearity in the log–log plot corresponds to the scale invariance associated with the observation of the phenomenon under consideration. Some 10 years after this methodology was published, Xiao [21] suggested a modified version of it by proposing two possible correction schemes. These can be achieved by implementing the end-to-end distance of the remaining residues and by performing $L(n)$ measurements starting from different C^{α} atoms, instead of only for the terminal C^{α}. These corrections increase the quality of the regression for high values of n, probably because it is only then the statistical nature of the process becomes more apparent. A detailed comparison of this genre of FD-based measures with the others that attempt to describe symmetry in protein backbone chain can be found in Bytautas et al. [22] and Aszódi and Taylor [23].

The FD magnitudes obtained from the first and second method, in the case of a self-avoiding walk (SAW) model, were 5/3 and 1.40, respectively [9, 10, 21]. It is known that, in three dimensions, FD will approach the topical two-dimensional Gaussian walk (or Θ solvent), whereas it will assume a magnitude 5/3 for SAW, which includes the excluded volume effect [24]. However, since the SAW does not take into account any attractive force, a deviation from idealized magnitude 5/3 is also expected, especially when there are interactions between monomers that might be far apart along the main-chain, but in close proximity within the folded protein [25]. Thus, more realistic polymers are likely to have a chain FD different from that of the theoretically predicted SAW magnitude, which is precisely what was observed from the obtained FD values. For more comprehensive discussion on FD-based description schemes for polymer collapse and polymer scaling in general, the reader is referred to previously published works [24, 26, 27]. However, it is important to remember that proteins are heteropolymers with a variable composition of 20 different amino acids and, additionally, according to the laws of polymer physics, proteins are very small (consisting of merely 50–250 amino acids in general). This is in stark contrast with procedures in the sphere of polymer studies because scaling relationships in polymer physics are (principally) derived for large homopolymers. Hence, one needs to be careful when attempting any possible extension of an algorithm and/or argument that is pertinent in the paradigm of polymer physics. (The reader is referred to Summary box 1 for a more detailed discussion on this topic).

Approach with residue (and atom) distribution

The second approach, which disregards any consideration of the connectivity profile of main-chain atoms, concentrates on the distribution of amino acid residues and atoms within the folded protein structures. Whether these residues and atoms are connected or not is not considered in this approach. This procedure (commonly called the 'mass fractal' calculation) studies the mass distributions within concentric spheres using, for example, the mass scales formulation $M \propto R^{\text{FD}}$ (where M is the total mass of all the protein atoms and R is the characteristic length scale). Thus, from studying the linear (scale invariant) portion of the slope of log M versus the log R plot, it is possible to determine the FD. Theoretically, one can accomplish this by counting the number of atoms within concentric spheres, taking any atom of the protein space as the center (see Fig. 3a for a schematic description). However, for most practical applications, the choice of the center of mass of the protein for the role of aforementioned center makes more sense. Although the generic name for the calculation of this genre of FD is mass fractals, this methodology can be used to calculate other pertinent properties associated with protein atoms. A larger mass-FD magnitude implies that within the protein interior the property under consideration is relatively more compact in terms of its distribution. The mass-FD magnitude varies as $2.00 < \text{mass-FD} < 3.00$. Proteins have an intrinsic self-similarity with respect to their compactness (atom packing). The mass-FD simply quantifies the symmetry of the scale-invariance associated with any protein property that is dependent on atom packing. It can also be extended to the level of residues, but not to the realm of organization of residues in secondary structures (α-helices are not composed of smaller helices). (The reader is referred to the Summary box 2 for a more detailed discussion on this topic).

This approach to FD calculation is rather new. Although reference to this approach can be found in Dewey's work

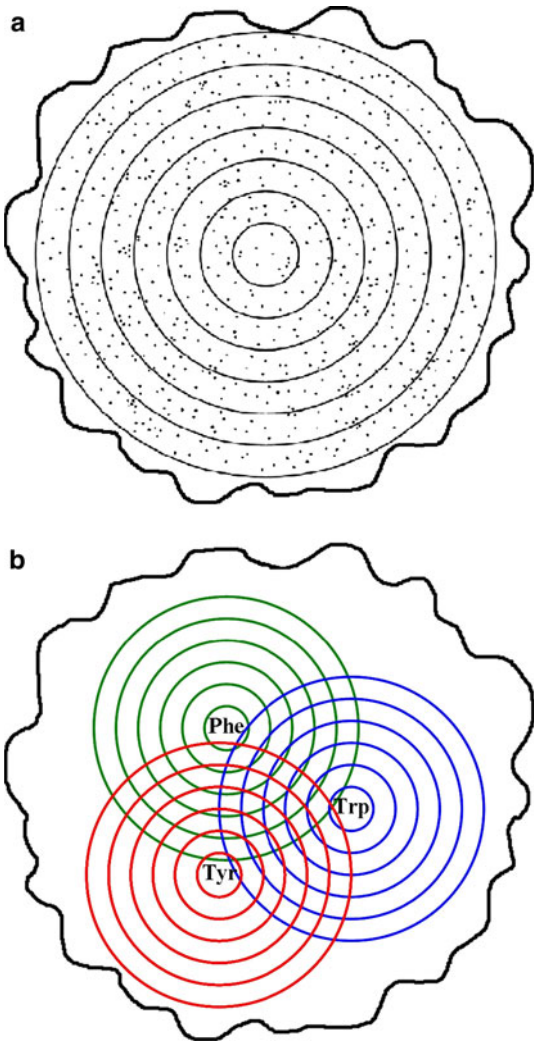

Fig. 3 **a** Schematic description of mass-fractal dimension calculation. The radial distribution of mass within the protein is obtained first, before studying the scaling of mass with respect to the radius. **b** Schematic description of the calculation of the correlation dimension (CD) in the context of CD amongst π-electron clouds; dependencies are identified before quantifying the extent of scaling of such dependencies with respect to radial distances

[28], the first rigorous attempt to describe proteins with this methodology occurred as recently as 2005 [29]. Thereafter, many studies [30–35] have utilized and extended the scope of this powerful construct to successfully unearth many forms of latent symmetries in the paradigm of protein interior interactions (these are discussed later in this article).

The mass-FD calculation, even though it presents a completely different standpoint of interior fractal study, is related to the approaches discussed above. The philosophy of the calculation of mass-FD can also be traced back to principles of polymer physics. The criterion of a 'good' solvent requires that the polymeric units have a preferential interaction with the solvent rather than with each other. As a consequence, the polymer acquires an extended shape and becomes what is called an 'excluded volume polymer'. In this case, the Flory constant υ_F assumes the magnitude 3/5. In an 'ideal' solvent (the so-called Θ solvent), the strength of the interaction of polymer with the solvent is identical to the strength of the interaction with itself. As a result, in an 'ideal' solvent, the polymer assumes a more compact shape than it does in a 'good' solvent, with $\upsilon_F = 1/2$. Polymer collapse takes place in the presence of a 'poor' solvent. In such a case, polymer–polymer interactions become more favorable than polymer–solvent interactions. Hence, the polymer forms a collapsed or a globule state with $\upsilon_F = 1/3$. This entire spectrum of possibilities can be elegantly described by $R_g \sim N\upsilon_F$, where R_g denotes the radius of gyration (ROG) of the polymer chain (a measure of its compactness), N stands for the number of units and Flory constant υ_F assumes different values depending on the solvent conditions. However, as $N \sim M$ and $M \sim R^{FD}$, we end up with $\upsilon_F = 1/FD$. Thus, the excluded volume polymer will have FD = 5/3. For an ideal polymer the FD magnitude will assume that of the topological dimension, namely, FD = 2; for a collapsed polymer, the FD will assume the magnitude of another topological dimension, namely FD = 3, whereby the polymer will assume a homogeneous space-filling nature. Neither $\upsilon_F = 1/2$ nor $\upsilon_F = 1/3$ is pertinent in the context of a functioning protein; in fact, they merely serve as the trivial limiting cases. Thus, while the relationship $2 <$ mass-FD < 3 holds for proteins, fledgling mass-FD studies can also be effortlessly associated with the established set of treatise of polymer physics. For a comprehensive review of polymer studies with FD measures, the reader is referred to a (classic) decade-old review [36]. (The reader is also referred to Summary box 1 for a more relevant discussion).

A note of caution, the discussion up to this point may give the reader the impression that the concept of FD is synonymous with that of ROG, but this is not the case. The reasons for their being different are discussed in detail later in this review. This discussion on ROG was included because most of the polymer physics works refer to ROG as a measure of compactness (instead of mass-FD).

Approach with the correlation dimension analysis

There are situations in which the researcher has a qualitative idea about the distribution of certain properties within the (protein) space, but the existence of a possible symmetry in their dependencies are not known. To obtain an objective view of the distribution of the dependencies, it is convenient to opt for the study of a different fractal dimension, namely, the 'correlation dimension' (CD). Informally, to determine the CD, the number of pair of

points between which a possible dependency might exist is first identified. Amongst these, the ones that lie within a sphere of radius r are counted, $C(r)$ (see Fig. 3b for a schematic description). By assuming a scaling $C(r) \sim r^{\mathrm{CD}}$, the Grassberger–Procaccia algorithm [37] is implemented to examine the possible presence of a linear stretch in the slope of a log–log plot of $C(r)$ versus r. Although a previous work [38] characterized the roughness of protein surfaces with CD, application of the same for protein interior studies occurred only very recently [39] where it was shown that the CD magnitude for ribosomal RNAs could successfully explain their structural characteristics (local helix formation and long-range tertiary interaction forming three-dimensional structures).

A categorical implementation of CD can be attempted by suitably describing each of the N monomers by some typical atom (say the C^{α} or centroid). Describing these atoms by points, the distance between two monomers i and j can be denoted as $|xi - xj|$. CD can then be formally defined as:

$$\mathrm{Corr}(r) = \frac{2}{N(N-1)} \sum_{i<j}^{N} \theta(r - |xi - xj|)$$

where, $\theta(x)$ is the Heaviside step function and $N(N-1)$ is a renormalization term [the presence of which implies that $\mathrm{Corr}(r)$ can be considered as the probability that any two residues are in contact at a cut-off distance r]. The term $\left[2\sum_{i<j}^{N} \theta(r - |xi - xj|)\right]$ describes symmetrical properties of the contact matrix.

{In fact, $\mathrm{Corr}(r)$ is an unbiased estimator of the correlation integral:

$$C(r) = \int \mathrm{d}\mu(x) \int \mathrm{d}\mu(y)\theta(r - |x - y|)$$

Both $\mathrm{Corr}(r)$ and $C(r)$, monotonically decrease to zero as $r \to 0$. If $C(r)$ decreases with a power law, namely, $C(r) \sim r^{D}$, then D (found from the slope) is called the correlation dimension of μ. Comprehensive theoretical discussion on this can be found elsewhere [40].

Finally, the dimension is defined by: $D = \lim_{r \to 0} \frac{\log \mathrm{Corr}(r)}{\log(r)}$. Details of the implementation of this algorithm along with a thorough discussion of several aspects of it can be found in previous publications [38, 39]. (The reader is referred to Summary box 2 for a more detailed discussion).

Approaches with renormalization group theory

As has already been mentioned, proteins are large macromolecules, and protein structures undergo large-scale fluctuations. In fact to ensure proper functioning, proteins often undergo large-amplitude movements [41, 42]. Since these fluctuations occur over all length scales, determination of a single characteristic scale parameter becomes extremely difficult (if not impossible). As a result, correlation functions between various biophysical properties within proteins typically demonstrate non-analytical behavior. More often than not, these correlation functions tend to follow fractional power law scaling [28] and depend on the dimensionality of the protein (under a given context), rather than on its microscopic characteristics.

Although the renormalization group theory (RGT) was originally applied to describe critical phenomena [43], the study of fractals and RGT are naturally linked because as in the case of fractal studies, the concept of scale invariance plays a major role in RGT. Briefly, the absence of an internal scale is one of the fundamental features of fractals. This can be rephrased by asserting that in fractals a great many number of scales coexist in a self-similar manner, which in turn reminds us of a comparable situation known from thermodynamic systems at the critical point. Divergence of correlation length at the critical point can be mathematically handled by assuming a composition of subsystems with some special coupling constant [44]. In such a case, each of the subsystems can be assumed to be built up of sub-subsystems with another coupling constant, and so on—which can reliably be called the self-similarity of protein interior organization. While FD-based studies attempt to quantify this self-similarity, the central idea of the RGT is that a change in the scale should lead to the same behavior apart from a renormalization of the coupling constant.

Two principal transformations are studied in the RGT paradigm, namely, the coarse graining and the scaling [45]. During the analysis of large-scale fluctuations of a protein, studies of its micro-scale perturbations might not always assume a great importance. Hence, in order to tackle the huge complexity associated with the fluctuations of proteins, researchers have come up with the efficient strategy of removing the length of the shortest interactions before the system is redefined by rescaling the length scale [28].

Since proteins exhibit self-similarity in their organization, this self-similarity can be captured with RGT and used to probe the various facets of (time-variant and context-dependent) interior dependencies. A categorical example of this can be found from the analysis of protein folding, which has been proved to be comparable to a phase transition [46, 47]. These problems are characterized by a multiplicity of scale lengths [one observes a large spectrum of correlations that correspond with the critical point, from short-range (among residues close by) to long-range (amongst residues situated far apart)]. This discussion of the scope of RGT implies that it can be suitable for a rigorous and objective description of such phenomena. Indeed, taking a cue from polymer physics studies [48], RGT has also been extensively applied in the realm of protein structure studies ([49–52] and references therein).

While Bohr and Wolynes [53] employed it for a protein folding study with contact energies, Chan and Dill [54] described its possible effectiveness for protein stability studies, and Coveney [55] used it for 'systematic coarse-graining' of a system of coupled nonlinear ordinary differential equations (see also relevant references therein) while studying protein interaction profiles. However, even after all these applications, the use of RGT has yet to become a common practice while attempting the scaling and coarse-graining operations in protein paradigm; as expressed in a recent work: 'rigorous mathematical procedures, such as RGT, have yet to be applied to the general definition of coarse-grain models' [56] while studying proteins.

Since FD- and RGT-based studies are closely related, there is an overlap between these approaches, and indeed many FD-based studies have explored their interface [57–59] with serious intent. It should not be forgotten, however, that these studies have been performed on a protein network paradigm and not on protein structure. In the sphere of protein interior studies, one merely finds a small number of FD studies that tangentially mention RGT [8, 17, 18] although there are two fantastic works [28, 60] in which the approaches and results of FD and RGT (in the context of polymer and proteins) are compared and contrasted in admirable detail.

Approach with the spectral dimension

Owing to the finite temperature effects, the constituent residues of proteins fluctuate around their native positions. An objective characterization of this dynamics is of pivotal importance when studying protein functions and is usually investigated through normal modes analysis (NMA) [61, 62]. A normal mode of an oscillating system is a pattern of motion in which all parts of the system move sinusoidally with the same frequency. The frequencies of the normal modes of a system are known as its natural frequencies or eigenfrequencies. The spectral dimension (d_S) governs the density of the low-frequency normal modes of a fractal or protein. It (d_S) is defined [63] according to the asymptotic behavior of the density of harmonic oscillations at low frequencies. To be precise, with the expression: $g(\omega) \sim \omega^{d_S-1}$, when $[\omega \to 0]$, where $g(\omega)$ denotes the density of modes with frequency (ω), and $g(\omega)d\omega$ denotes the number of modes in the frequency range $[\omega, \omega + d\omega]$. Hence, d_S can undoubtedly be identified as a marker directly related to the vibrational dynamics of a fractal.

Although an old study [64] on the interpretation of fractals provides a hint of this concept, its formalization can be found first an article by Burioni and Cassi [65] and subsequently in two other published studies [66, 67]. The applicability of this construct increases significantly when it was based upon the generalization [66] of the Landau–Peierls instability criterion [68], which provides a connection between the spectral dimension (d_S) and FD. The number of amino acids along the protein backbone N has been established recently [30]. This relationship is given by:

$$\frac{2}{d_S} + \frac{1}{\mathrm{FD}} = 1 + \frac{b}{\ln(N)}$$

where the parameter b shows a weak dependence on temperature and interaction parameters. The impact of this derivation can be easily ascertained by observing the varied application of it (notably in [68, 69]) in problems across protein structure studies. For a detailed biophysical understanding of the interplay of fractal and spectral dimensions, the reader is referred to the article of Reuveni [3].

Qualitatively, the spectral dimension describes the effective connectedness within protein structures, of which a large magnitude corresponds to high topological connectedness (the reader is referred to Summary box 2 for a more detailed discussion). This complements the mass-FD magnitude, which typically quantifies how completely a protein property fills the protein space. The spectral dimension magnitude coincides with the Euclidean dimension in the case of lattices, but in general, d_S assumes non-integer values of between 1 and 3.

Results obtained with fractal-based investigations

Fractal-based procedures can extract valuable information about the latent symmetry in terms of the arrangement of atoms and residues. They can also be indispensable in searching out patterns from a spectrum of nonlinear, time-dependent and context-dependent interactions amongst the interior biophysical properties. Thus, not surprisingly, these algorithms have been applied (meaningfully) in a wide range of cases. Due to the aperiodic arrangement of the amino acids, conformational states of proteins consist of many sub-states with nearly the same energy [70]. Experimental evidence indicates that each sub-state of the protein has in itself a large number of sub-states, and the potential energy function is statistically self-similar, having the same form on many different scales [51]. Here we present the principal findings from a wide variety of FD-based studies.

FD-based protein conformation studies

Fractal dimension is a reliable indicator of protein conformation because it provides a quantitative measure of the

degree to which a structure (and any property associated with the structure) fills the space. Based on this recognition, as far back as 1982, Stapleton and his co-workers [71] showed that changes in protein structure under solvent conditions could conveniently be monitored by FD. In fact, their studies [17, 18, 71, 72] reported for the first time that the geometry of the carbon backbone determined from X-ray diffraction data and the vibrational dynamics of proteins as measured by Raman scattering are both fractals. The results of their work proved that FD determined from the structure correlates (almost) perfectly with that determined from the dynamics. Subsequently, in 1984, another group took this question up and worked further on the Stapleton finding that the FD of the polymer (ω) coincides with the d_S that governs the phonon density of the states $g(\omega)$ at small frequencies as $g(\omega) \sim \omega^{d_S-1}$. However, their interpretation [73] of fracton was challenged [74] soon after (on the ground that fracton dimension only considers the scalar excitations, while the phonons are vector excitations), and the exponent in the Stapleton finding was related to the spectral dimension. For an in-depth account of these concepts, interested readers are referred to earlier publications ([18, 73–75] and references therein), to appreciate their pivotal significance in the context of contemporary theoretical soft condensed matter research.

FD-based ion-channel kinetics studies

Liebovitch and co-workers [76–78] proposed a fractal model of ion channel kinetics in 1987. The fact that their model was more consistent with the conformational dynamics of proteins helped in establishing that fractal models can be accurate and reliable when conformational transitions between states that consist of a hierarchy of substates are taken into consideration. Liebovitch et al.'s interpretation of the results was radical because it was based on the notion that dynamic processes in proteins occur with many different correlation times. By adjusting the parameters of the model, the observed gating behavior could be described over a wide time range. The immediate impact of this study on their then contemporary studies can be understood by comparing the (varying) approaches and interpretations suggested a number of these studies [79–82] with those that were attempting to describe the same genre of results with other tools. Subsequent works by Liebovitch [83, 84] on the analysis of patch clamp recordings of the sequence of open and closed times of cell membrane ion channels were able to establish a new trend in ion channel kinetics studies, which asserted that ion channel proteins have many conformational states (connected by large number of pathways) of nearly equal energy minima. These states were shown not to be independent but, rather, linked by physical mechanisms that result in the observed fractal scaling. The impact of these findings can be appreciated from the profile of succeeding studies [85–89] on fractal kinetics from 1991 onwards. For example, in a recent work [89] it has been proved that patterns in spiking activity in suprachiasmatic nucleus neurons can be explained with the fractal point process model, which implies the presence of self-similarity in the profile of spike trains, proving thereby that no characteristic time scale dominates the dynamics of the spiking process.

FD-based attempts to relate protein structure and dynamics

It has long been known that protein dynamics involves a broad range of interconnected events occurring on various time scales [90]. It is therefore hardly surprising that protein systems exhibit nonexponential behavior. The fractal approach facilitates the development of models of protein dynamical processes. There is a wealth of nonexponential rate processes in biophysics, and fractal models can be effective in describing such processes [91]. Although some dynamical models [92, 93] had been developed to describe the vast array of nonexponential processes, discriminating between them was a difficult (yet important) problem. Fractal models have a unique appeal they allow a structural parameter, the FD, to (automatically) be related to a dynamic parameter, such as d_S, the spectral dimension. Dewey [27, 91, 94] explored this inherent direct connection between structure and dynamics in a series of studies. To study the emergence and characteristics of these nonexponential behaviors involved in the conformational change of proteins and protein structures in general, the reader is referred to one of Dewey's articles [28]. Recent applications of this approach can be found in a number of publications [95–97].

Studies in fractal kinetics

The discussion on fractal studies of ion channel kinetics and nonexponential rate processes falls under the broad category of 'fractal kinetics' study. It was Kopelmann [98] who suggested that the general phenomenon of logarithmically decreasing reaction rates should be collectively considered under the umbrella term of 'fractal kinetics'. His assertions were proven experimentally under a number of conditions [99]. Spatial and/or energetic heterogeneity of the medium, or non-randomness of the reactant distribution in low dimensions, has been suggested as (possible) mechanisms underlying fractal kinetics. Kopelman [99] showed how surface diffusion-controlled reactions, which occur on clusters or islands, can be expected to exhibit anomalous and fractal-like kinetics. Mechanisms of fractal kinetics exhibit anomalous reaction orders and time-

dependent (for example, binding) rate coefficients, being therefore advantageous. Investigations [100] (with theoretical considerations raised by statistical methods of random walk on fractal structures) have proven that fractal kinetics is not due to the number of binding sites on each protein molecule but, rather, is a reflection of the fractal properties of enzymes (and proteins in general) themselves. Since the roughness of protein surfaces follows fractal scaling [101] and FD is a global, statistical property, insensitive to time-dependent tiny fluctuations, the results in [100] can be easily understood. The same study had (notably) established that non-integral dimensions of the Hill coefficient (used to describe the allosteric effects of proteins and enzymes) are direct consequences of fractal properties of proteins. Another consequence of this inherent fractal nature is demonstrated by the power-law dependence of a correlation function at certain coordinates and the power-law dependence of the correlation function on the surface of analyte-receptor complexes over time [95]. The original Kopelman algorithm was given strong theoretical support when it was shown that fractal kinetics owes its emergence to spatial self-organization of the reactants induced by the compact properties of diffusion [102]. Being an extremely significant platform, countless applications of this scheme have been attempted over the years, with a number of these being quite important [103–107]. The interested reader might also benefit from studying the work of Turner et al. [103] who carried out an in-depth comparison of various pertinent schemes.

A discussion on fractal kinetics cannot be complete without a note on fractal time studies. However, this area falls outside the purview of this review. The interested reader can find an excellent discussion on chemical kinetics and fractal time and their pertinence in the realm of protein interaction studies elsewhere ([108], pp 158–163 of [28]).

FD-based results on protein structure

The recent trend of FD-based studies indicates a strong inclination towards discovering protein structural invariants. Since mass-FD is a measure of the compactness, it can alternatively be used as an effective measure of the extent of torsion that the extended polypeptide undergoes during folding: the stronger the torsion of the systems, the larger it's mass-FD, and vice versa. Therefore, FD reflects, to a large extent, the structural information of proteins and is itself a reliable construct to study the extent of folding. Enright and Leitner [29] formalized this intuitive notion by systematically calculating the mass-FD of a set of 200 proteins. They found the magnitude to be 2.489 ± 0.172, which was consistent with those obtained from dispersion relations and the anomalous subdiffusion computed for the same set of proteins [109], which were (in turn) related to the theory of mass fractal dimension proposed by Alexander and Orbach [63]. A large body of incisive studies performed during the last 3–4 years has zoomed in on two aspects of Enright–Leitner's finding, i.e., the influence of the magnitude of the mass-FD on protein dynamics and energy flow and the non-compactness of the folded primary structure. The findings from these studies were unexpected. Here we outline a number of them.

The similarity between the geometry of a globular protein and a percolation cluster was pointed out by Liang and Dill [7]. However, a (masterly) follow-up study [110] revealed that energy flow through the vibrational states of proteins is anisotropic and that this process can be reliably described by the theory of transport on a percolation cluster in terms of the dimension of the cluster and the density of vibrational states. Although the precise functional roles of the energy propagation channels could not be identified, the researchers proposed the possibility of relating these to the fast cooling of the proteins during the reaction, directed energy transport to speed up the reaction, communication signals and allostery.

Adopting the fractal point of view made it possible to analyze protein topology and protein dynamics within the same framework. This advantageous fact was utilized wonderfully in a recent work by De Leeuw et al. [2] in which a universal equation of state for protein topology was derived. Although the marginal stability criterion of proteins had been discussed in an earlier study [111], De Leeuw et al. [2] proved that most proteins 'exploit' the Landau–Peierls instability [112] to attain large-amplitude vibrations. The authors insisted that this was the key mechanism by which proteins perform large-amplitude conformational changes to ensure proper function while maintaining an invariant native fold. Furthermore (perhaps most importantly), they proved that the majority of proteins in the Protein Data Bank (PDB) exist in a marginally stable thermodynamic state, namely a state that is close to the edge of unfolding. These represent exciting new findings on the state of existence of proteins, and they could never have been brought to the fore without efficient and accurate implementation of FD-based concepts.

The horizon in this area has recently opened up yet further when the answer to a basic question, namely, 'why, although proteins seem to vibrate under a complex energetic landscape thanks to their highly involved dynamics, the same energetic landscape appear harmonic near the minima?' was answered. This was achieved [113] by challenging the very validity of the harmonic approximation concerned, and the study was able to establish that anomalous vibrational dynamics in proteins is a manifestation of a fractal-like native state structure. To obtain an even more inclusive picture of this aspect of protein structure studies, one can complement the aforementioned

[113] assertion with that from another recent study [114] in which the latter authors were able to identify a possible mechanism, namely, the small fractal dimension of 'first passage networks', that accounts for reliable folding in proteins with rugged energy landscapes. On a related note, it is useful to know that a work on protein folding that is too slow to be simulated directly, asserted that the fractal Smoluchowski equation could reliably model protein folding by subdiffusion [115]. In this way, this methodology could achieve what the standard paradigm of either diffusion (the usual Smoluchowski equation) or normal-diffusion continuous time random walk of a single order parameter under thermodynamic influences would have had difficulty achieving.

Another equally unexpected set of results [31] was presented when the degrees of non-compactness of mass, hydrophobicity and polarizability within proteins were quantified across four basic structural classes. This study revealed that barring all-α class, all of the major structural classes of proteins have an amount of unused hydrophobicity left in them. The presence of unused hydrophobicity in the bacterial type 3A cellulose-binding domain protein '1nbc' (all-beta domains) is shown in Figure 4. The amount of unused hydrophobicity was observed to be greater in thermophilic proteins than in their (structurally aligned) mesophilic counterparts. The all-β proteins (thermophilic, mesophilic alike) were identified with a maximum amount of unused hydrophobicity, while all-α proteins were found to possess minimum polarizability. These results proved that the origin of α-helices are possibly not hydrophobic but electrostatic, whereas β-sheets are predominantly hydrophobic in nature.

Staying with non-compactness, but approaching the questions from a different perspective, the backbone of the ribosome was examined in another study [38] by employing correlation dimension (discussed in section Approach with correlation dimension analysis). The results of this study revealed that capacity dimensions of rRNAs are less than their embedding dimension, implying the presence of some empty spaces in rRNA backbones. Going back to the principles of polymer physics in the context of proteins, Hong and Jinzhi [35] derived a unified formula for the scaling exponent of proteins under different solvent conditions. Since this formula was obtained by considering the balance between the excluded volume effect and elastic interactions among monomers, one can expect it to be the focus of a lot of attention in future studies. Hong and Jinzhi's results showed that the scaling exponent is closely

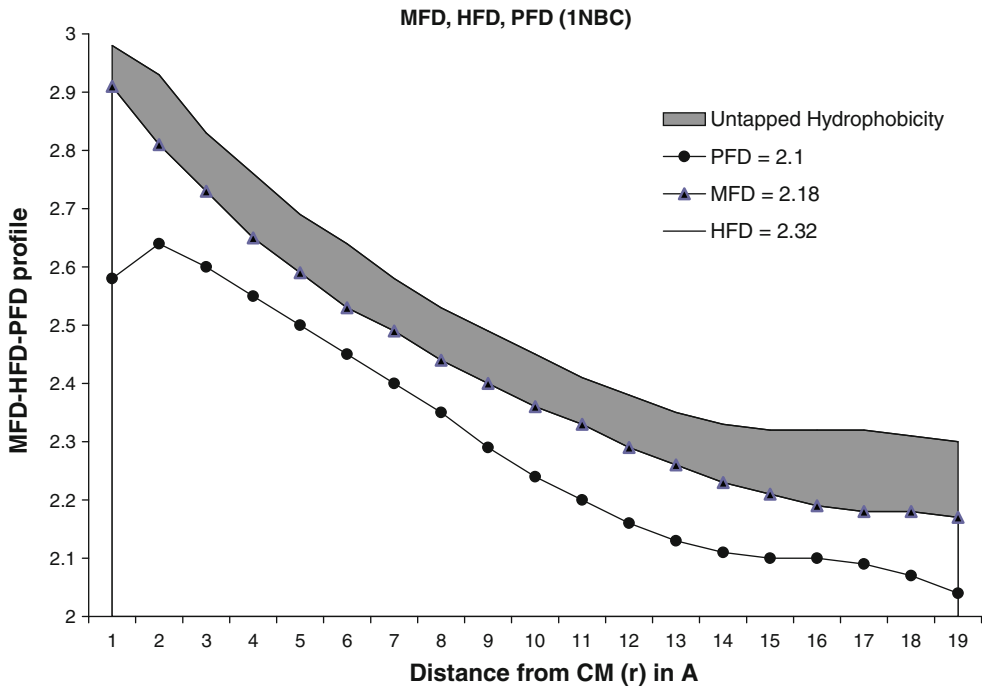

Fig. 4 Profiles of the mass-fractal (*MFD*), hydrophobicity-fractal (*HFD*) and polarizability fractal (*PFD*) dimensions for the protein (chosen randomly) 'bacterial type 3a cellulose-binding domain' (PDB id.: 1NBC; no. of residues 155, radius of gyration 8.53) are plotted simultaneously. Stretches in these profiles that are linear and parallel to the abscissa (i.e. radial distance from the center of mass of the protein) denote the extent to which these profiles are scale-invariant, with variations in radial distance. Ordinate magnitudes corresponding to these stretches provide the value of MFD, HFD and PFD for the protein. The [+(HFD − MFD] parameter quantifies the magnitude of unused hydrophobicity left in the protein. In the present case, ([+(HFD − MFD]) is given by [+(HFD − MFD] = [(2.32 − 2.18)] space-filling unit = [+0.14] space-filling unit

related to the FD of a protein's structure at the equilibrium state. Hidden dependencies between protein structural class-specific FD magnitudes and kinetic–thermodynamic parameters (the folding and unfolding rate, folding–unfolding free energy) were recently studied [39]. The results of this study confirmed the dependence of FD values on the fold type and on the location and connectivity of the secondary structures. However, the authors left a cautionary note by reporting that the presence of turns tends to increase the values of FD, irrespective of the structural classes. In an effort to describe accessible surface area, the number of amino acids and the radius of gyration with a FD-centric protein view, a recent work on protein (non)compactness [116] found that compactness of a protein might well be described in terms of packing between random spheres in percolation threshold and crumpled wires.

Although the description of a single protein molecule as a contact network of amino acid residues is not new, a recent study [117] has investigated the properties of the shortest path length between the number of nodes by employing the network topological dimension, the fractal dimension and the spectral dimension. The authors proved conclusively that residue contact networks of proteins in native structures are not small world networks but fractal networks.

Gaining new knowledge of the protein interior with FD

While various successful applications of FD-based constructs in probing protein interiors are described in the section Results obtained with fractal-based investigations, much is yet to be done in this area. Just to emphasize this very point, that is, how as yet-unexplored questions can be rigorously investigated through the innovative use of FD constructs, we present here a glimpse of correlation-dimension-based investigations of protein structural parameters. Since the underlying symmetries can be studied in short, medium and long-range interactions between various residues with an invariant framework, correlation dimension (CD)-based studies on the protein interior can prove to be extremely useful. The scope and resolution of the obtained results underscore the point the application of this methodology in a series of elaborate and systematic studies will provide extensive information on numerous (elusive) dependencies between biophysical and/or biochemical invariants within proteins.

Implementation of CD-based constructs

Correlation dimension (as explained in section Approach with correlation dimension analysis) is different than the other FD constructs, because it attempts to investigate the self-similarity in distributions between correlated properties (or dependencies) which arise out of particular dispositions of atoms or residues. In other words, instead of studying the self-similarity in distribution of atoms or residues—something that the mass-fractal analyses [explained in section Approach with residue (and atom) distribution] performs—here, one examines the self-similar characteristics of correlations amongst the properties that come to being due to any given distribution of atoms or residues. A correlation dimension-based analysis of key biophysical properties necessary for protein stabilization is shown in Fig. 5. While some of these are quantifiable through other possible means, others, say the possible self-similarity in the distribution of π-electron clouds of the aromatic amino acids, will be difficult to obtain through other procedures. While in the section CD-based investigation of local and global dependencies amongst peptide dipoles units we describe how SCOP (structural classification of proteins [118])-class centric inferences can be obtained with CD-based constructs, in the section CD-based investigation of dependency distribution amongst hydrophobic residues, charged residues and residues with π-electron clouds we describe the utility of this methodology in quantifying the extent of dependencies amongst properties of individual residues. The section CD-based investigation of local and global dependencies amongst peptide dipoles units therefore presents the results of general nature, whereas the section CD-based investigation of dependency distribution amongst hydrophobic residues, charged residues and residues with π-electron clouds presents the results at the resolution of particular proteins. Instead of merely concentrating on characterization of self-similarity in electrostatic interactions, the general problem of protein stability was addressed by separately investigating the self-similarity in correlations among peptide dipoles, charged amino acids, aromatic amino acids and hydrophobic amino acids. Conserved patterns derived from the results of these four independent investigations provided a unique scheme by which to quantify protein stability at a general (SCOP fold and SCOP class) level.

CD-based investigation of local and global dependencies amongst peptide-unit dipoles

Why is it advantageous to resort to CD while studying protein electrostatics?

Many quantitative analyses have attempted to describe the delicate balance of forces in the complex charge subsystem and hard-to-model dielectrics of the protein interior. Proposed simulation strategies to study this problem range from continuum dielectric methods [119] to an explicit approach for dealing with polarizability [120]. In other

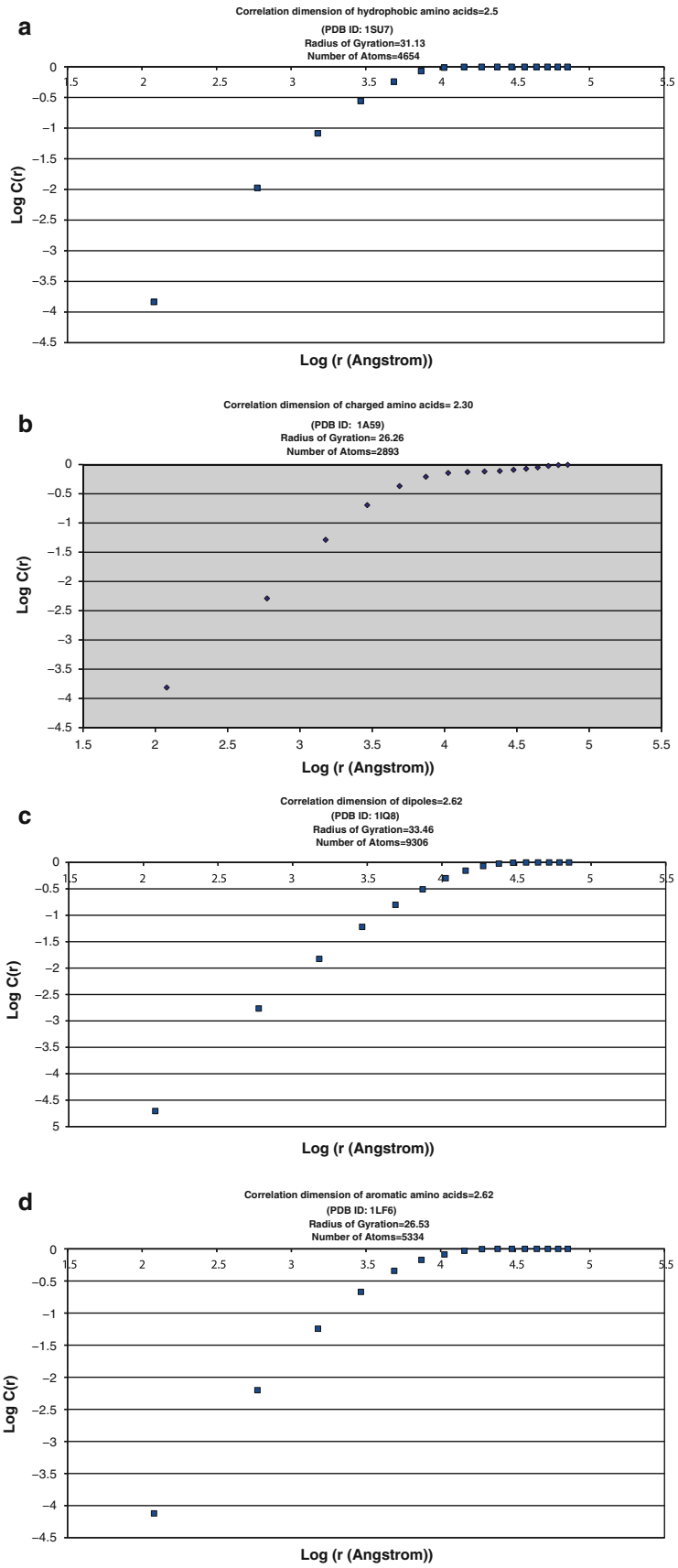

Fig. 5 Denoting the spatial correlation between two residues separated by distance r (alternatively, the probability of two residues to be in contact at a cut-off distance r) by $C(r)$, limit of $[\log(C(r))/\log(r)]$ was calculated when $r \to 0$. This magnitude quantified the correlation dimension for each of the cases. In other words, from the slope of each of the four curves, the corresponding correlation dimension was calculated. Details of the methodology and background for each of these figures can be found in the sections Approach with correlation dimension analysis and FD-based attempts to relate protein structure and dynamics. **a** CD of hydrophobic amino acid, **b** CD of charged amino acid, **c** CD of dipoles and **d** CD of aromatic amino acid for selected proteins are shown

words, starting from the Tanford–Kirkwood dielectric cavity scheme, the analytical solution of the Poisson–Boltzmann equation, the non-linear numerical finite difference [121], boundary element algorithms [122] and, finally, the sophisticated empirical generalized Born solutions [123] have all been tried. However, the explicit treatment of polarizability, even at its modest linear response level, involves a high-computational price and detailed know-how of molecular modeling practices. Thus, the immediate need of many protein scientists, i.e. an easy-to-understand and easy-to-calculate measure to describe protein electrostatics, is not always addressed by the aforementioned procedures. In this context, the use of CD-based analyses of the protein interior may be immensely beneficial. Since charge–charge and peptide dipole interactions have been identified to play important roles in the stabilization of proteins [124] and in determining the native fold [125], we demonstrate here how CD-based investigations on these can bring valuable information to the fore using an invariant simple framework.

Calculation of CD to quantify dipole distribution

Although the peptide group is uncharged at a normal pH, its double-bonded resonance form accounts for a dipole moment of approximately 3.6 Debye, directed from oxygen of C=O, to hydrogen of N–H bond. In the case of certain secondary structures, such as the α-helix, these dipole moments align with each other to produce a significant resultant dipole moment, which in turn may account for non-trivial contributions to protein stability. The electrostatic field resulting from the arrangement of helix dipoles has been proposed as an important factor contributing to both helix stability and specificity [126]. Åqvist [127] and coworkers have demonstrated that this effect is short-ranged, and its influence is confined largely to individual backbone dipoles localized within the first and last turns of the helix, resulting in a formal positive charge at the helix N terminus and formal negative charge at the C terminus. However, the global picture regarding protein interior electrostatics, as a function of backbone conformation of the native structure, does not necessarily emerge from that. Since dipoles in an α-helix are electrostatically aligned nearly parallel to the axis of the helix [128], such a global picture can be obtained straightaway by calculating the general dependency profile between dipoles originating from the numerous mutual spatial dependencies between the peptide-unit dipoles across the entire backbone. The mere vectorial summation to calculate the resultant dipole vector will fall short of quantifying this aspect of dipole distribution in protein structures.

Hence, choosing the standard threshold for considering a pair of residues to be in long-range contact [129, 130], namely a maximal distance of 8 Å between their centroids [alternatively, the distance between C^{β} atoms C^{α} for glycines) can be chosen; the results do not vary appreciably (unpublished data)] and resorting to the formulae provided in section Approach with correlation dimension analysis, the CD were calculated for every non-redundant protein structure (taken from the PDB [131]) within all of the SCOP folds with a statistically significant number of non-redundant protein structures. All of the non-hydrogen atoms were considered while calculating the centroid of an amino acid. The set of formulae mentioned in the section Approach with correlation dimension analysis was used to calculate the CD. The reader is referred to Theiler [132] for details of the methodology. After obtaining the CD for each of the proteins belonging to a particular SCOP fold, average of obtained CDs for a particular SCOP fold was subsequently calculated to draw biological inferences at the SCOP-class level. While Tables 1 and 2 contain the list of 15 SCOP folds with the maximum and minimum magnitudes of CD between peptide dipoles, respectively, the entire list can be found in the Electronic Supplementary Material 1.

Brief discussion on peptide-unit dipole distribution within SCOP folds

Most of the studies on peptide-unit dipoles [126, 128] have concentrated on the nature and magnitude of the (resultant) dipole moments, especially in the context of α-helices. Results obtained with CD, on the other hand, have quantified a different aspect of peptide-unit dipoles in proteins, namely, quantification of the prevalent symmetry in dependencies amongst the spatial arrangement of the dipoles across proteins from all seven SCOP classes. Somewhat unexpectedly, the results revealed that peptide-unit dipoles in α/β class of proteins are, in general, more correlated to each other than the peptide-unit dipoles in proteins belonging to all-α class proteins. Of 15 SCOP folds with a maximum CD amongst peptide-unit dipoles (Table 1), only one entry ('α/α toroid', with the maximum CD) belongs to the all-α class of proteins, nine belong to the α/β class of proteins, four belong to the $\alpha + \beta$ class of proteins, and one is from multidomain proteins. Hence, although peptide-unit dipoles in α-helices are electrostatically aligned nearly parallel to the axis of the helix [128] and although their resultant effect in helix stability and specificity [126] is well-known, with respect to ensuring maximum dipole–dipole interactions, the α/β structures appear to be better optimized. Then again, in the particular case of fold 'α/α toroid', the closed circular arrangement of array of α-hairpins presents perfect positioning for the peptide-unit dipoles to correlate with each other, which explains α/α toroid's having the highest CD for

Table 1 List of 15 SCOP folds with maximum correlation dimensions between peptide-unit dipoles

SCOP class	SCOP fold	Magnitude of correlation dimension
All-α proteins	α/α Toroid	2.490
α/β proteins	PLP-dependent transferase-like	2.469
Multi-domain protein	β-Lactamase/transpeptidase-like	2.463
α/β proteins	α/β-Hydrolases	2.452
$\alpha + \beta$ proteins	Protein kinase-like (PK-like)	2.432
α/β proteins	TIM β/α-barrel	2.428
α/β proteins	Ribokinase-like	2.426
α/β proteins	Nucleotide-diphospho-sugar transferases	2.407
α/β proteins	Periplasmic binding protein-like II	2.379
α/β proteins	*S*-Adenosyl-L-methionine-dependent methyl transferases	2.368
$\alpha + \beta$ proteins	Class II aaRS and biotin synthetases	2.363
$\alpha + \beta$ proteins	Cysteine proteinases	2.342
α/β proteins	Phosphorylase/hydrolase-like	2.337
α/β proteins	HAD-like	2.327
$\alpha + \beta$ proteins	Ntn hydrolase-like	2.325

SCOP, Structural classification of proteins

Table 2 List of 15 SCOP folds with minimum correlation dimensions between peptide-unit dipoles

SCOP class	SCOP fold	Magnitude of correlation dimension
Peptides	Conotoxins	0.539
Peptides	Transmembrane helical fragments	0.933
Coiled coil proteins	Parallel coiled coil	1.095
Designed proteins	Zinc finger design	1.125
Small proteins	Knottins (small inhibitors, toxins, lectins)	1.252
Small proteins	β-β-α Zinc fingers	1.309
Small proteins	Glucocorticoid receptor-like (DNA-binding domain)	1.360
Coiled coil proteins	Stalk segment of viral fusion proteins	1.431
Small proteins	Rubredoxin-like	1.512
All-α proteins	RuvA C terminal domain like	1.628
All-α proteins	Long-α hairpin	1.675
All-β proteins	Glycosyl hydrolase domain	1.730
All-β proteins	SH3-like barrel	1.768
All-α proteins	λ-repressor like DNA binding domains	1.834
All-α proteins	DNA RNA binding 3 helical bundle	1.854

peptide-unit dipoles. Notably no structural domain belonging to SCOP class all-β was found in the list of folds with a maximum CD between peptide-unit dipoles.

The assertion of better inherent correlations between dipoles in α/β proteins is vindicated by Table 3 and by results shown in Table 2 too. Amongst the 15 SCOP folds with a minimum CD between peptide-unit dipoles, not a single entry was found in α/β class of proteins. Four entries from the 'small proteins' class underline the fact that in this SCOP fold, the correlation (and dependencies) between peptide-unit dipoles on each other is extremely less. Interestingly, four folds from the all-α class are found to have the least spatial correlations amongst peptide-unit dipoles which, supported by the data in Table 1, contradicts the popular notion that dipole–dipole interactions within all-α proteins are the best. On the other hand,

Table 3 Correlation dimension between peptide-unit dipoles across four major SCOP classes

SCOP class	Dipole moment correlation dimension
α/β proteins	0.958 ± 0.252 (Highest)
$\alpha + \beta$ proteins	0.725 ± 0.220
All-α proteins	0.629 ± 0.241
All-β proteins	0.624 ± 0.258 (Least)

Data are presented as the mean $\pm$ standard deviation (SD)

Table 3 results show unambiguously that the spatial correlation (and therefore, dependency) between peptide dipoles in all-α and all-β class of proteins is almost same. This result is, however, not unexpected because CD does not take into account the directionality of dipoles, but merely attempts to quantify their spatial correlations.

CD amongst charged amino acids across all seven SCOP classes

Considering His, Arg, Lys, Asp and Glu as the charged amino acids, we carried out an analysis on the non-redundant structures from all seven SCOP classes with the same methodology as used in the previous case. Table 4 shows 15 SCOP folds with maximum values of CD amongst charged amino acids, and Table 5 shows 15 SCOP folds with minimum magnitudes of CD amongst charged amino acids. Interestingly, eight out of top 15 CDs between charged amino acids were observed in proteins from the α/β class (Table 4). This is striking similar to the results obtained from the SCOP-class-wide peptide-unit dipole correlation pattern (Table 1), where nine of the top 15 CD magnitudes were observed in proteins from the α/β class. The optimized electrostatic environment of α/β proteins can alternatively be assessed from their (conspicuous) absence from Table 5. Although it was expected that the proteins from the SCOP-class 'small proteins' and 'designed proteins' might have a minimal CD among the (aforementioned) five charged residues (Table 5), it was unexpected to observe minimal CDs among those five charged residues even among the proteins from three all-α folds ('RuvA C terminal domain like', 'λ-repressor like DNA binding domains' and 'SAM domain like'). Furthermore, since these are exactly the same folds with minimal CD amongst peptide-unit dipoles (Table 2), it may be indirectly inferred that the electrostatic environment within the proteins belonging to these domains is minimally conducive.

CD amongst aromatic amino acids across all seven SCOP classes

It is known [133] that Tyr, Trp and Phe possess the 'π-system'. While many studies have been conducted on characterizing cation–π interactions, π–π stacking, among others, the symmetry in spatial correlation between π-electron clouds is far from being understood. Considering all of the (non-hydrogen) atoms belonging to aromatic rings of Tyr, Trp and Phe, we carried out an analysis on the non-redundant structures from all the seven SCOP classes, using the same methodology as in the previous case. Table 6 shows 15 SCOP folds with maximum values of CD amongst π-electron clouds of the aromatic amino acids, whereas Table 7 shows 15 SCOP folds with minimum magnitudes of CD amongst the same aromatic amino acids. Conforming to the trends observed in the distribution of peptide-unit dipoles and charged amino acids, six out of top 15 CDs between aromatic amino acids were observed in proteins from the α/β class (Table 6). Aromatic amino acids from proteins belonging to the $\alpha + \beta$ class feature in the same list for four of the top 15 cases. The marked absence of proteins belonging to either of these SCOP classes from Table 7 implies the existence of conducive electrostatic environments within α/β and $\alpha + \beta$ proteins in

Table 4 List of SCOP folds with a maximum correlation dimension amongst the charged amino acids (His, Arg, Lys, Asp and Glu)

SCOP class	SCOP fold	Magnitude of correlation dimension
$\alpha + \beta$ proteins	Class II aaRS and biotin synthetases	2.324
$\alpha + \beta$ proteins	Protein kinase-like	2.322
$\alpha + \beta$ proteins	Lysozyme-like	2.275
α/β proteins	PLP-dependent transferase-like	2.273
All-α proteins	α/α toroid	2.261
α/β proteins	HAD-like	2.258
Multidomain proteins	β-Lactamase/transpeptidase-like	2.253
$\alpha + \beta$ proteins	Zincin-like	2.251
$\alpha + \beta$ proteins	Cysteine proteinases	2.241
$\alpha + \beta$ proteins	ATP-grasp	2.240
α/β proteins	Nucleotide-diphospho-sugar transferases	2.239
α/β proteins	TIM β/α-barrel	2.238
α/β proteins	*S*-Adenosyl-L-methionine-dependent methyltransferases	2.234
α/β proteins	α/β-Hydrolases	2.212
α/β proteins	Ribokinase-like	2.208

Table 5 List of SCOP folds with a minimum correlation dimension amongst the charged amino acids (His, Arg, Lys, Asp and Glu)

SCOP class	SCOP fold	Magnitude of correlation dimension
Peptides	Conotoxins	0.920
Peptides	Transmembrane helical fragments	1.250
Coiled coil proteins	Stalk segment of viral fusion proteins	1.488
Small proteins	Glucocorticoid receptor-like (DNA-binding domain)	1.606
Coiled coil proteins	Parallel coiled coil	1.628
Small proteins	Knottins (small inhibitors, toxins, lectins)	1.711
Designed proteins	Zinc finger design	1.729
All-β proteins	Glycosyl hydrolase domain	1.769
All-α proteins	RuvA C terminal domain like	1.863
Small proteins	β-β-α Zinc fingers	1.887
Small proteins	Rubredoxin-like	1.929
All-α proteins	Long-α hairpin	1.949
All-β proteins	Immunoglobulin like β-sandwich	1.969
$\alpha + \beta$ proteins	FKBP-like	1.973
All-α proteins	λ-Repressor-like DNA binding domains	1.980

Table 6 List of SCOP folds with a maximum correlation dimension amongst the aromatic amino acids

SCOP class	SCOP fold	Magnitude of correlation dimension
Multidomain proteins	β-Lactamase/transpeptidase-like	2.531
All-α proteins	α/α Toroid	2.497
All-β proteins	β-Trefoil	2.424
All-β proteins	GroES like	2.389
All-β proteins	Trypsin-like serine proteases	2.387
$\alpha + \beta$ proteins	Protein kinase-like	2.386
α/β proteins	TIM β/α-barrel	2.381
α/β proteins	α/β-Hydrolases	2.348
$\alpha + \beta$ proteins	Cysteine proteinases	2.341
α/β proteins	Nucleotide-diphospho-sugar transferases	2.338
$\alpha + \beta$ proteins	ATP-grasp	2.334
α/β proteins	Adenine nucleotide α-hydrolase-like	2.333
α/β proteins	Ribokinase-like	2.328
α/β proteins	PLP-dependent transferase-like	2.325
$\alpha + \beta$ proteins	Ntn hydrolase-like	2.311

general, resulting in interactions between π-electron clouds of Tyr, Trp and Phe being maximized. On an unexpected note, interactions between aromatic amino acids of proteins belonging to SCOP class all-α were observed to be minimal amongst all seven SCOP classes, which is demonstrated by the presence of six all-α folds in Table 7.

CD amongst hydrophobic amino acids across all seven SCOP classes

Analysis of self-similarity in interactions amongst hydrophobic amino acids was performed by quantifying their dependencies (or correlations) on each other. This can be achieved by calculating the CDs amongst Gly, Ala, Val, Ile, Leu, Met and Cys for proteins belonging to the seven SCOP classes. The observed pattern, namely, that more than half of the SCOP folds with maximum values of CD amongst the hydrophobic amino acids belong to the α/β class (Table 8), point definitely to the superior stability profile of proteins belonging to the α/β structural class in general. Such an assertion of better stability is further vindicated by a marked absence of α/β folds in Table 9, which shows 15 SCOP folds with minimum values of CD amongst the hydrophobic amino acids. Although

Table 7 List of SCOP folds with a minimum correlation dimension amongst the aromatic amino acids

SCOP class	SCOP fold	Magnitude of correlation dimension
Peptides	Conotoxins	0.369
Designed proteins	Zinc finger design	1.512
Small proteins	Rubredoxin-like	1.199
All-α proteins	RuvA C terminal domain like	1.276
Coiled coil proteins	Parallel coiled coil	1.320
All-α proteins	λ-repressor like DNA binding domains	1.408
Small proteins	Glucocorticoid receptor-like (DNA-binding domain)	1.414
All-α proteins	EF hand like	1.445
All-α proteins	Spectrin repeat like	1.498
Peptides	Transmembrane helical fragments	1.506
All-β proteins	SH3-like barrel	1.517
Small proteins	Knottins (small inhibitors, toxins, lectins)	1.544
All-α proteins	SAM domain-like	1.603
Small proteins	β-β-α Zinc fingers	1.605
All-α proteins	Four helical up and down bundle	1.616

Table 8 List of SCOP folds with a maximum correlation dimension amongst the hydrophobic amino acids

SCOP class	SCOP fold	Magnitude of correlation dimension
All-α proteins	α/α Toroid	2.358
α/β proteins	α/β-Hydrolases	2.315
α/β proteins	PLP-dependent transferase-like	2.309
Multidomain proteins	β-Lactamase/transpeptidase-like	2.295
α/β proteins	TIM β/α-barrel	2.287
α/β proteins	Ribokinase-like	2.285
α + β proteins	Protein kinase-like	2.267
α/β proteins	Nucleotide-diphospho-sugar transferases	2.259
α + β proteins	Cysteine proteinases	2.242
α + β proteins	Class II aaRS and biotin synthetases	2.236
α/β proteins	*S*-Adenosyl-L-methionine-dependent methyltransferases	2.223
All-β proteins	Concanavalin A like lectins glucanases	2.201
α/β proteins	Phosphorylase/hydrolase-like	2.196
All-β proteins	Acid proteases	2.192
α/β proteins	Periplasmic binding protein-like II	2.188

'peptides', 'small proteins' and 'designed proteins' were expected to have negligible hydrophobic interaction profiles, the presence of four all-α folds ('RuvA C terminal domain like', 'long-α hairpin', 'λ-repressor like DNA binding domains' and 'SAM domain like') and two all-β folds ('SH3 like barrel' and 'glycosyl hydrolase domain') in Table 9 indicate that it will be simplistic to expect an enhanced nature of hydrophobic interactions amongst proteins from many classes belonging to the all-α and all-β SCOP classes.

General inferences about protein stability from SCOP-class-wide separate analyses of CD between charged residues, the peptide dipole, aromatic amino acids and hydrophobic amino acids

The CD-based analyses conducted here involved the study of self-similarity according to three aspects of protein electrostatics as well as a separate examination of self-similarity in hydrophobic interactions. Although studies of protein stability are numerous and multifaceted, quantitative

Table 9 List of SCOP folds with a minimum correlation dimension amongst the hydrophobic amino acids

SCOP class	SCOP fold	Magnitude of correlation dimension
Peptides	Conotoxins	0.497
Peptides	Transmembrane helical fragments	1.045
Small proteins	Knottins (small inhibitors, toxins, lectins)	1.103
Designed proteins	Zinc finger design	1.118
Coiled coil proteins	Parallel coiled coil	1.157
Small proteins	β-β-α Zinc fingers	1.191
Small proteins	Glucocorticoid receptor-like (DNA-binding domain)	1.305
Small proteins	Rubredoxin-like	1.357
All-α proteins	RuvA C terminal domain like	1.379
Coiled coil proteins	Stalk segment of viral fusion proteins	1.413
All-α proteins	Long-α hairpin	1.498
All-α proteins	λ-Repressor like DNA binding domains	1.582
All-β proteins	SH3-like barrel	1.587
All-β proteins	Glycosyl hydrolase domain	1.611
All-α proteins	SAM domain-like	1.614

investigations of self-similarity in interactions that influence protein stability on a completely global level are difficult to find. Even from the limited analyses conducted here, certain significant information on the latent features of protein stability could be unearthed, including a detailed characterization of the electrostatic environment within protein structures belonging to various SCOP folds across all seven SCOP classes. For example, we found that in the four tables showing the names of SCOP folds with maximum CD magnitudes (Tables 1, 4, 6 and 8, each comprising 15 SCOP folds), the names of nine SCOP folds appear consistently. Proteins belonging to these nine SCOP folds, namely, α/α toroid (all-α), α/β-hydrolases (α/β), PLP-dependent transferase-like (α/β), TIM β/α-barrel (α/β), β-lactamase/transpeptidase-like (multidomain proteins), cysteine proteinases ($\alpha + \beta$), nucleotide-diphospho-sugar transferases (α/β), protein kinase-like ($\alpha + \beta$) and ribokinase-like (α/β), can therefore be considered to be those having the maximum stability in general. Such a conserved profile of symmetry in stability-enhancing factors can further be observed in proteins belonging to folds '*S*-adenosyl-L-methionine-dependent methyl transferases'(α/β) and Class II aaRS and biotin synthetases ($\alpha + \beta$) because they appear in three of the aforementioned four tables. The fact that some of these SCOP folds with maximal CD magnitudes are amongst the most evolutionarily conserved folds (say, TIM β/α-barrel, α/β-hydrolases, etc.) is therefore not surprising.

On a diametrically oppsing scenario, one finds an even more conserved profile in the lack of stability by observing the consistency in the occurrence of the names of SCOP folds with minimum CD magnitudes (Tables 2, 5, 7 and 9, each presenting 15 SCOP folds). To elaborate, ten SCOP folds appear consistently in all four tables, while four other SCOP folds appear in three tables. Proteins belonging to these SCOP folds, namely, conotoxins (peptides), transmembrane helical fragments (peptides), knottins (small peptides), zinc finger design (designed proteins), parallel coiled coil (coiled coil proteins), β-β-α zinc fingers (small proteins), glucocorticoid receptor-like (small proteins), rubredoxin-like (small proteins), RuvA C terminal domain-like (all-α) and λ-repressor-like DNA binding domains (all-α), are therefore expected to have a very low stability profile. Amongst these, the fold conotoxins (peptides) is conspicuous in its securing a minimum CD magnitude in all four aspects of the current investigation, indicating the least possible stability for proteins belonging to this fold. Four other SCOP folds are observed to be conserved across three of the aforementioned tables, i.e. SH3-like barrel (all-β), long-α hairpin (all-α), glycosyl hydrolase domain (all-β) and 'stalk segment of viral fusion proteins' (coiled coil proteins), indicating also a lack of stability of proteins belonging to these folds. Concentrating solely on the electrostatic environment, one can identify two folds (conotoxins and transmembrane helical fragments) of class 'peptides' as having the least conducive electrostatic milieu since they are found to register the least CD magnitudes in correlations among the peptide dipoles and, separately, among the charged amino acids (His, Arg, Lys, Asp and Glu).

These results clearly indicate that many more latent and general patterns of a reported nature can be unearthed by systematic studies of protein interior properties with CD-based analyses.

CD-based investigation of dependency distribution amongst hydrophobic residues, charged residues and residues with π-electron clouds

The reasons for the usefulness of CD-based analyses in studying protein electrostatics have been discussed in sections Why is it advantageous to resort to CD while studying protein electrostatics? and General inferences of protein stability from SCOP-class-wide separate analyses of CD between charged residues, peptide-unit dipole, aromatic amino acids and hydrophobic amino acids. SCOP-class centric (top–down) analysis to demonstrate the technique has been provided in sections Calculation of CD to quantify dipole distribution and CD amongst hydrophobic amino acids across all seven SCOP classes. Here we show the utility of the same analysis based on the amino acid centric (bottom–up) framework. Considering the centroids (instead of C^{α}-atoms) of amino acids as true indicators of positions, the magnitude of CD (found from the slope of the log–log plot) between predominantly hydrophobic amino acids (Fig. 5a) within the protein 'carbon monoxide dehydrogenase (PDB id: 1su7) was calculated. (As in the case of previous calculations, the set of formulae mentioned in section Approach with correlation dimension analysis was used. The reader is referred to a previous publication [132] for details of the methodology.) Similarly, considering His, Arg, Lys, Asp and Glu as the charged amino acids, the correlation dimension between charged amino acids was calculated for the protein 'cold-active citrate synthase' (PDB id: 1a59) (Fig. 5b). To quantify the self-similarity prevalent in the dipole moment distribution of a complex multidomain protein 'archaeosine tRNA-guanine transglycosylase' (PDB id: 1iq8), we calculated the CD (shown in Fig. 5c). Equally interesting was the study of the CD between π-electron clouds of Tyr, Trp and Phe in 'bacterial glucoamylase' (PDB id: 1lf6) (Fig. 5d). Although Fig. 5 shows the results obtained from individual protein structures, it is easy to note that comprehensive scale-up of these on a statistical scale can unearth hidden dependencies between pertinent entities embodying biophysical or biochemical properties.

Problems with FD and necessary precautions while working with FD

There are a number of inherent problems associated with FD-based toolkits, and whoever is interested in exploring the scope of utilizing FD-based measures should be aware of these. For example, as a consequence of the scale-invariance of fractals, the functions describing the latter can never be smooth. This is a significant characteristic because it implies that the normal practice of expanding a function to the first order and then to approach the linearized problem will not be compatible with FD-based constructs. The scale-invariance of fractals ensures that one cannot find a scale where functions become smooth enough to linearlize. Due to this innate non-analytic nature, fractals have undefined derivatives, making the application of common principles of differential calculus to them absolutely useless. Thus, the very property that makes fractals useful for describing irregular natural shapes (and phenomena) are in themselves problematic.

On another note, the incongruity between the (idealistic) geometrical requirements of a fractal object and the (real-life) statistical character of the obtained results might seem a bit disconcerting for beginners. While the strict definition of FD demands an 'infinite scale reduction', this cannot be expected in calculations with real-systems (say, proteins) because the size of a protein is restricted—i.e. it cannot span until infinity. Thus, the FD of a protein property can only be measured within a range of scale. On the other hand, for many protein properties, the magnitude of FD might show a dependence on the number of residues or atoms—which, again, contradicts the idealistic expectations. However, experience suggests that for a system with a statistically significant number of residues (or atoms in some cases), many of these problems disappear—at least when calculating the FD within a certain scaling range and not across the entire possible scaling range. If, however, the aforementioned dependencies are still found (in other words, if no scale-invariant zones are found to describe the distribution of the property concerned), it is possible to suspect that the distribution of the property is not self-similar in the first place.

These problems can be translated into certain (common) categorical problems:

1. For example, there are non-trivial differences between proteins and 'classical collapsed polymers' (pp. 41–44 of [28]). Hence, some of the theoretical constructs of polymer physics cannot be applied 'directly' in the realm of protein structure studies. Similarly, while it is known that energy surfaces of proteins are rough hypersurfaces in high-dimensional configuration spaces with an immense number of local minima [70] and that proteins share this feature with other complex systems, such as spin glasses, glass-forming liquids, macromolecular melts, among others [134], one should not forget that compared to the aforementioned complex systems, a protein is a relatively small system. Thus, a blind application of FD-based procedures pertinent for spin-glass studies might not be judicious in the realm of proteins.
2. Furthermore, one should always be aware of the statistical nature of FD-based measures. This implies

that they can be useful for studying a property from a 'top–down' perspective but not from a 'bottom–up' standpoint, in cases where the atoms are distinguishable. Thus, one needs to be careful while studying the 'local' profile of certain properties with FD measures. Unfortunately, ascertaining the limit of this 'local'-ness in terms of the precise number of (interacting) parameters is primarily dependent upon experience. However, a reasonable approach can be to consider a statistically significant number of such interactions. Defining any of the fractal exponents with five to ten atoms makes no sense. On the other hand, problems with aspects of an interesting parameter averaging out due to the innate constraints of FD-based methodologies is not uncommon (see, for example, [38]). Drawing from that, one can assert that probing the protein interior with fractal exponents for some 'nano' or 'microscopic' properties (or emergence of it) might be extremely risky. Similarly, one should be aware of the possibility that a system might show statistical self-similarity only when the X and Y-axes describing it are magnified by different scales. These are called 'self-affine fractals' and while they are not exactly common, they are also not rare in the paradigm of protein studies [135].

3. FD-based analyses often provide results in (somewhat) abstract form that are difficult to be explained fully from our present state of knowledge of biophysics and biochemistry. (Examples of such results can be found in many recent and old works [117, 136, 137]). Thus, one should be patient in deciphering the abstract nature of fractal results.

Summary and outlook

Various approaches to calculate the FD of proteins have been discussed in this review. The results obtained from the studies presented complement each other. Even a cursory look at the various paper repositories will demonstrate the rapid growth in the number of published studies during the last 3 years that have attempted to explore the protein interior with fractal exponents. This body of research collectively presents the contour of a promising multifaceted technique that can honestly and reliably probe an object as complex as protein. Indeed, a lot has been achieved using the FD; however, one can observe that many published studies end with proposed new studies to take the obtained information to completeness. For example, even after the (pioneering) work of Leitner (discussed in section FD-based results on protein structure) [110], one needs to relate the complete information of energy transport channels within individual proteins before associating them to their functional roles. Similarly, studies need to be performed with the 'universal equation of state' (discussed in sections Approach with the spectral dimension and FD-based results on protein structure) [2, 30] to fully understand the biophysical and biochemical implications of protein's existence at 'the edge of unfolding' before attempting to use it for practical protein engineering applications. Likewise, in another recent work [31] (discussed in section FD-based results on protein structure), non-trivial correlations between polarizability-FD and hydrophobicity-FD in all-α, α/β and (especially) $\alpha + \beta$ proteins have been reported. Do these imply a small yet significant dependence of the dielectric constant on the hydrophobicity content of a protein (and vice versa)? Similarly, one needs to explore the uniqueness of distributions of structural properties in all-β proteins to answer why neither their mass nor hydrophobicity distribution show any dependence on their polarizability distributions. In this context, perhaps an even more challenging problem will be to ascertain exactly how many atoms are needed to observe the emergence of mass-FD or hydrophobicity-FD or polarizability-FD? Since, as a recent work indicates, the emergence of protein's biochemical and biophysical properties with its statistical nature might well be mesoscopic in nature [138] and since FD in the context of protein structure studies essentially quantifies the statistical self-similarity, the evaluation of the exact number of atoms to observe the emergence of interior FD parameters will be immensely beneficial.

None of the questions and none of the points discussed in the paragraph above are simple. However, many of the unsolved problems in the realm of FD-based protein interior studies can be identified with far more (surprisingly) simple questions. For example, as mentioned in the sections Approach with residue (and atom) distribution and Approaches with backbone connectivity, the FD of the protein backbone is independent of the secondary structural content of the protein, which implies that secondary structural classes can reliably be modeled as Hamiltonian walks [28]. (Hamiltonian walk is a lattice walk in which every point is visited once and only once, with no intersecting paths; it can be used to model polymers that show compact local and global scaling, such as 'crumpled globules' [139]). Crumpled globules tend to imply an absence of knots in the protein structures [28]. However, knots do occur in proteins and not entirely rarely [140, 141]. How does one resolve this contradiction from an objective general perspective?

Referring to an even simpler (?) case, we concentrate on amino acid frequency studies. The distribution of pairs of amino acids is not a linear combination of frequencies of occurrences of the constituent individual amino acids [142], and the distribution of a given residue along a

polypeptide chain is fractal in nature [137]. However, the latter study also reports a negative correlation between the probabilities of occurrence of various residues and their FDs for which, to the best of our knowledge, no general satisfactory answer has been proposed. Similarly, although a recent study [117] establishes that the amino acid residue network in the protein native structure forms a fractal—and not small-world—network, no general biophysical cause could be attributed to this. At the same time, however, the authors' explanation of the reasons underlying the percolated nature of the amino acid residue-contact network suggests the possibility of incorporating a recent assertion [2] that the majority of proteins in the PDB exist in a marginally stable thermodynamic state, namely, a state that is close to the edge of unfolding. A comprehensive study involving two of these from the perspective of molecular evolution might provide us with unique perception of protein structures.

Hence, while a lot of unknown, unexpected and deep information on (nonlinear) dependencies between protein structural parameters can be unearthed with FD-based probing, one needs to be careful while using the FD tools and interpreting the results obtained through their application. It should never be forgotten that apart from their prominent importance in biological functions, proteins are qualified subjects for studying the dynamics of complex systems. The FD-based investigations of protein interior organization can serve as a unique platform for interactive studies between physicists and biologists. However, to end on a cautionary note, although FD-based investigations of protein properties are insightful and although they can often reveal unexpected information, FD-based measures are not panacea. Therefore, one needs to be aware of the intrinsic limitations of these methodologies and at the same time take necessary precautions (as outlined before) while implementing them.

Acknowledgments One of the authors, Anirban, would like to thank the COE-DBT scheme, Govt. of India, for supporting him during the tenure of this work. He would also like to thank Charudatta Navare for various technical help and Professor Paul Bourke for permitting him to use the self-similarity figures.

Summary box 1: Comparing and contrasting scaling laws in polymers and proteins

An elaborate comparative discussion on fractal and spectral properties of proteins and polymers is far beyond the scope of this review. Indeed, a large body of excellent publications [18, 24, 26, 36, 54, 139, 151, 152] can be found on this topic. This summary box attempts to merely augment the information provided on this topic in the section Results obtained with fractal-based investigations.

X-ray crystallographers and polymer scientists have long asserted that proteins are very compact collapsed polymers. Their view, in short, derives from the fact that many first-generation mass fractal dimension (MFD) results suggested that the dimension of the contour of the protein backbone is slightly less than 3. Polymers below the θ-point are referred to as collapsed polymers and have a FD of 3. Statistically, therefore, proteins were considered as collapsed polymers. Recent results [29, 31], however, have shown that MFD magnitudes in proteins are significantly smaller than those expected from a completely compact 3-D collapsed polymer. The reports of inhomogeneous packing [6] in the protein interior and presence of cavities therein [7] support the recent findings.

Hence, a thorough knowledge of fractal studies in the realm of polymers might not be indispensable while studying the same in the paradigm of proteins. However, studying the fractal properties of protein primary structures from the polymer perspective helps us to understand protein fractal properties in the general sense. For proteins in physiological conditions, the scaling law $R_g \propto N^{2/5}$ (where R_g = ROG, N = no. of residues) was observed [148, 149].

According to Flory's original theory [pp. 109–129 of [139], pp. 519–538 of [150], pp. 24–32 of [151], pp. 97–104 of [152], the radius of gyration of a homopolymer at the equilibrium state depends on the chain length through a scaling law $R_g \propto N^{\nu}$, where the exponent ν depends on the solvent conditions. Under poor solvent conditions, one finds $\nu = 1/3$, whereas under good solvent conditions, one finds $\nu = 3/5$. However, proteins are heteropolymers made up of 20 different kinds of amino acids, each of which can be either hydrophobic or hydrophilic to a varying extent. If, in a hypothetical case, all of the residues are found to be hydrophobic, the situation will correspond to poor solvent conditions. Here, the protein will be highly compressed by solvent pressure, and the magnitude of the scaling exponent ν will be 1/3. On the other equally hypothetical case, if all residues are found to be hydrophilic, the situation will correspond to good solvent conditions. Here, the protein will be extended with $\nu = 3/5$, to form as many hydrogen bonds with the surrounding water molecules as possible. In reality, the solvent condition for a general natural protein under physiological conditions stays between a good and poor solvent condition. Hence, native structures of proteins show a scaling law with $\nu = 2/5$. Remarkably, $\nu = 2/5$ holds across proteins belonging to the four major protein structural classes (all-α, all-β, $\alpha + \beta$, α/β) [35], indicating that the scaling process is perhaps independent of protein's secondary and tertiary organization.

Since proteins are long-chain molecules, can we neglect the differences between characteristics of each amino acids—at least while considering the statistical properties

of the large-scale structures, such as the radius of gyration and the scaling exponent? Such a treatment can allow a direct application of Flory's original idea for homopolymers in the realm of proteins. While this seems logical to expect, especially if one modifies the interaction forms used in Flory's work to incorporate the characteristics of proteins, the issue might not be a straightforward one. The scaling exponent magnitude $\nu = 2/5$ helps in comparing protein and polymer behaviors. The fact that scaling exponent ν satisfies $1/3 < \nu < 3/5$ indicates that folded proteins are more compact than polymers in a good solvent, but wobblier than highly compressed polymers in a poor solvent. Henceforth, structural specificity of natural proteins appears to be distinct from that of usual homopolymers [35]. Moreover, owing to multiple constraints in the process of geometrical packing of several domains, large proteins may rarely acquire idealistic spherical shapes. One therefore needs to be careful when extending the principles of polymer physics in the paradigm of proteins.

Summary box-2: Various types of FDs and their interrelationships

Proteins are heteropolymers with a variable composition of 20 amino acids. Natively folded proteins can be characterized by (at least) three types of Hausdorff dimensions: the mass fractal dimension (MFD), the spectral dimension (SD) and the correlation dimension (CD). None of these dimensions attempt to quantify particular details of individual residues or atoms (say, how the side-chain torsion angle of a particular residue will change when the solvent pH changes, etc.). Rather, all of them, from three different perspectives, attempt to acquire collective knowledge of the global patterns that influence protein structural organization.

The MFD describes the spatial distribution of mass within the protein through the scaling relationship $M(r) \sim r^{\mathrm{MFD}}$, where $M(r)$ is the mass enclosed in a sphere of radius r, MFD being the mass fractal dimension itself. It is important to note that the word 'mass' is metaphorical here; one can measure other properties (say, hydrophobicity, polarizability etc.) as well with MFD as long as specific magnitudes of these properties can be assigned to a relevant point in protein-space (say, an atom or a residue). MFD, therefore, can be a generic term that quantifies how completely a property fills the protein space. Many compactness studies thus resort to the calculation of MFD. Although the magnitude of MFD correlates well with the radius of gyration of proteins, [29, 31], possible reasons for calculating the MFD (instead of radius of gyration) can be found from a recent work [31].

The SD quantifies the density of low-frequency vibrational normal modes via the scaling relation: $g(w) \propto w^{(\mathrm{SD}-1)}$, where $g(w)dw$ is the number of modes in the frequency range $[w, w + dw]$[143]. Theoretically, one can consider SD to be an informal analogue of Euclidean dimension applied to disordered structures in the context of dynamic processes [67]. In the case of the lattices, the magnitudes of the SD and Euclidean dimensions do not differ. In general, however, SD assumes non-integer magnitudes between 1 and 3. Far more extensively studied than MFD and CD, SD indicates the extent of effective connectedness of geometrical structures on a large scale because large values of SD correspond to high topological connectedness. Starting from quantifying harmonic oscillations and anomalous density of vibrational modes in proteins, SD helps in retrieving information about diffusion, phase transitions and electrical conductivity [67].

The CD is a measure of the dimensionality of the space occupied by a set of random points. Since it describes the spatial relationship of one point to all other points within a set of data, a measure of correlation between these points, with respect to a certain property, can be obtained with CD. Algorithmically, CD captures how a given number of points within a distance L scales with L. In comparison to protein-structural knowledge obtained through the use of MFD and SD, the same gathered from applying CD is minuscule. the sections FD-based attempts to relate protein structure and dynamics and Problems with FD, necessary precautions while working with FD of this review, however, present a CD-based analysis of protein interior properties.

To understand how these three dimensions are related to each other and how they complement each other in terms of acquiring protein structural information, we need to appreciate their physical implications. The MFD quantifies the extent of the space-filling nature of certain properties on protein-space. Different points [coordinates of atoms or coordinates of C^{α} (or else, centroid) of residues] may have different magnitudes of the aforementioned property assigned to them. The distribution of any property (as displayed by different magnitudes of it across different points) will be a function of the structural organization of proteins. Conversely, the manner in which various points, each endowed with a different magnitude of aforementioned property, interact with each other may influence protein structural organization. In other words, the possible self-similar symmetries in the interactions between protein points endowed with a property can provide unique information about the importance of the distribution of this property in ensuring organizational standards for protein structure under consideration. The information obtained through MFD, however, is insufficient for this purpose. Instead, it is the CD which measures the extent of

self-similarity in the dependencies (or possibility of interactions) amongst points in protein-space that are endowed with different magnitudes of the same property.

Since MFD can quantify the compactness of a protein, it can also quantify the extent of folding. In other words, MFD can measure the extent of torsion present in a folded protein, with the greater the torsion of the protein backbone, the larger its MFD. Hence, MFD can be utilized to retrieve much related information on the extent of folding, and all of this can be achieved from an invariant platform. However, another peculiarity of protein structure is that its constituent residues fluctuate about their native positions in a temperature-dependent manner. Amongst many other effects (enzyme catalysis, function and activity of proteins in general), residue fluctuation influences the thermal properties of proteins, such as the heat capacities. It has been found that protein heat capacity at low temperature follows the relationship $C_V(T) \propto T^{SD}$. In fact, in general, while one may use SD to quantify the large-scale topology of a geometric structure, SD can be used to quantify low-energy fluctuations of a given protein structure. As a consequence, one can derive an instability criterion for proteins, banking only on the topological considerations.

The relationship between MFD and SD can be understood most easily while studying energy flow on percolation cluster. The mean square displacement of a vibrational excitation on a fractal object was found to vary as $:R^2 \sim t^{\gamma}$, where $\gamma = \text{SD/MFD}$. In fact, information obtained with MFD and SD complement each other; while MFD provides information on the arrangement of the atoms of the protein, the SD, although lacking direct kinetic information, relates to the dynamics of the system and the density of the vibrational states. For somewhat more technical (yet highly illuminating) discussions on the inter-relationship between MFD and SD, interested readers can immensely benefit from information provided in various publications (see references [64, 143–147, 154]).

References

1. Karplus M (2000) Aspects of protein reaction dynamics: deviations from simple behavior. J Phys Chem B 104:11–27
2. De Leeuw M, Reuveni S, Klafter J, Granek R (2009) Coexistence of flexibility and stability of proteins: an equation of state. PLoS One 4(10):e7296
3. Reuveni S (2008) Proteins: unraveling universality in a realm of specificity. PhD thesis. Tel Aviv University, Tel Aviv
4. Goetze T, Brickmann J (1992) Self similarity of protein surfaces. Biophys J 61:109–118
5. Tissen J, Fraaije J, Drenth J, Berendsen H (1994) Mesoscopic theories for protein crystal growth. Acta Cryst D 50:569–571
6. Banerji A, Ghosh I (2009) A new computational model to study mass inhomogeneity and hydrophobicity inhomogeneity in proteins. Eur Biophys J 38:577–587
7. Liang J, Dill KA (2001) Are proteins well-packed? Biophys J 81:751–766
8. Havlin S, Ben-Avraham D (1982) New approach to self-avoiding walks as a critical phenomenon. J Phys A 15:L321–L328
9. Havlin S, Ben-Avraham D (1982) Fractal dimensionality of polymer chains. J Phys A 15:L311–L316
10. Havlin S, Ben-Avraham D (1982) New method of analysing self-avoiding walks in four dimensions. J Phys A 15:L317–L320
11. Mandelbrot BB (1982) The fractal geometry of nature. W. H. Freeman and Co, San Francisco
12. Hausdorff F (1919) Dimension und ¨ausseres Mass. Math Ann 79:157–179
13. Barnsley M (1988) Fractals everywhere. Academic Press, San Diego
14. Falconer K (1990) Fractal geometry: mathematical foundations and applications. Wiley, New York
15. Meakin P (1998) Fractals scaling, and growth far from equilibrium. Cambridge University Press, Cambridge
16. Isogai Y, Itoh T (1984) Fractal analysis of tertiary structure of protein molecule. J Phys Soc Jpn 53:2162
17. Wagner GC, Colvin JT, Allen JP, Stapleton HJ (1985) Fractal models of protein structure, dynamics, and magnetic relaxation. J Am Chem Soc 107:20
18. Colvin JT, Stapleton HJ (1985) Fractal and spectral dimensions of biopolymer chains: solvent studies of electron spin relaxation rates in myoglobin azide. J Chem Phys 82:10
19. Tanford C (1961) Physical chemistry of macromolecules. Wiley, New York
20. Wang CX, Shi YY, Huang FH (1990) Fractal study of tertiary structure of proteins. Phys Rev A 41:7043–7048
21. Xiao Y (1994) Comment on fractal study of tertiary structure of proteins. Phys Rev E 46:6
22. Bytautas L, Klein DJ, Randic M, Pisanski T (2000) Foldedness in linear polymers: a difference between graphical and Euclidean distances, DIMACS. Ser Discr Math Theor Comput Sci 51:39–61
23. Aszódi A, Taylor WR (1993) Connection topology of proteins. Bioinformatics 9:523–529
24. Gennes P (1996) Scaling concepts in polymer physics. Cornell University Press, Ithaca
25. Elber R (1989) Fractal analysis of protein. In: Avnir D (ed) The fractal approach to heterogeneous chemistry. Wiley, New York
26. Dewey TG (1993) Protein structure and polymer collapse. J Chem Phys 98:2250–2257
27. Dewey TG (1995) Fractal dimensions of proteins: what are we learning? Het Chem Rev 2:91–101
28. Dewey TG (1997) Fractals in molecular biophysics. Oxford University Press, New York
29. Enright MB, Leitner DM (2005) Mass fractal dimension and the compactness of proteins. Phys Rev E 71:011912
30. Reuveni S, Granek R, Klafter J (2008) Proteins: coexistence of stability and flexibility. Phys Rev Lett 100:208101
31. Banerji A, Ghosh I (2009) Revisiting the myths of protein interior: studying proteins with mass-fractal hydrophobicity-fractal and polarizability-fractal dimensions. PLoS One 4(10):e7361
32. Moret MA, Miranda JG, Nogueira E Jr, Santana MC, Zebende GF (2005) Self-similarity and protein chains. Phys Rev E 71:012901
33. Lee CY (2006) Mass fractal dimension of the ribosome and implication of its dynamic characteristics. Phys Rev E 73:042901
34. Figueirêdo PH, Moret MA, Nogueira E Jr, Coutinho S (2008) Dihedral-angle Gaussian distribution driving protein folding. Phys A 387:2019–2024

35. Hong L, Jinzhi L (2009) Scaling law for the radius of gyration of proteins and its dependence on hydrophobicity. J Polym Sci Part B 47:207–214
36. Novikov VU, Kozlov GV (2000) Structure and properties of polymers in terms of the fractal approach. Russ Chem Rev 69:523–549
37. Grassberger P, Procaccia I (1983) Measuring the strangeness of strange attractors. Phys D 9:183–208
38. Lee CY (2008) Self-similarity of biopolymer backbones in the ribosome. Phys A 387:4871–4880
39. Tejera E, Machadoa A, Rebelo I, Nieto-Villar J (2009) Fractal protein structure revisited: topological kinetic and thermodynamic relationships. Phys A 388:4600–4608
40. Takens F (1985) On the numerical determination of the dimension of an attractor. In: Braaksma B, Broer H, Takens F (eds) Lecture notes in mathematics, vol 1125. Springer, Berlin, pp 99–106
41. Steinbach PJ, Ansari A, Berendzen J, Braunstein D, Chu K, Cowen BR, Ehrenstein D, Frauenfelder H, Johnson JB, Lamb DC (1991) Ligand binding to heme proteins: connection between dynamics and function. Biochemistry 30:3988–4001
42. Rasmussen BF, Stock AM, Ringe D, Petsko GA (1992) Crystalline ribonuclease A loses function below the dynamical transition at 220 K. Nature 357:423–424
43. Wilson KG (1975) The renormalization group: critical phenomena and the Kondo problem. Rev Mod Phys 47:773
44. Wilson KG (1979) Problems in physics with many scales of length. Sci Am 241:140–157
45. Goldenfeld N (1992) Lectures on phase transitions and the renormalization group. Addison-Wesley, Reading
46. Goldstein RA, Luthey-Schulten ZA, Wolynes PG (1992) Protein tertiary structure recognition using optimized Hamiltonians with local interactions. Proc Natl Acad Sci USA 89:9029–9033
47. Goldstein RA, Luthey-Schulten ZA, Wolynes PG (1992) Optimal protein-folding codes from spin-glass theory. Proc Natl Acad Sci USA 89:4918–4922
48. Family F (1982) Direct renormalization group study of loops in polymer. Phys Lett 92A:341–344
49. Pierri CL, Grassi AD, Turi A (2008) Lattices for ab initio protein structure prediction. Protein Struct Funct Bioinf 73:351–361
50. Böhm G (1991) Protein folding and deterministic chaos: limits of protein folding simulations and calculations. Chaos Solitons Fractals 1:375–382
51. Li HQ, Chen SH, Zhao HM (1990) Fractal structure and conformational entropy of protein chain. Int J Biol Macromol 12:374–378
52. Nonnenmacher TF (1989) Fractal scaling mechanisms in biomembranes. Eur Biophys J 16:375–379
53. Bohr HG, Wolynes PG (1992) Initial events of protein folding from an information-processing viewpoint. Phys Rev A 46:5242–5248
54. Chan HS, Dill KA (1991) Polymer principles in protein structure and stability. Annu Rev Biophys Chem 20:447–490
55. Coveney PV, Fowler PW (2005) Modelling biological complexity: a physical scientist's perspective. J R Soc Interf 2:267–280
56. Heath AP, Kavraki LE, Clementi C (2007) From coarse-grain to all-atom: toward multiscale analysis of protein landscapes. Proteins 68:646–661
57. Song C, Havlin S, Makse HA (2006) Origins of fractality in the growth of complex networks. Nat Phys 2:275–281
58. Kitsak M, Havlin S, Paul G, Riccaboni M, Pammolli F, Stanley HE (2007) Betweenness centrality of fractal and nonfractal scale-free model networks and tests on real networks. Phys Rev E 75:056115
59. Rozenfeld HD, Song C, Makse HA (2010) The small world-fractal transition in complex networks through renormalization group. Phys Rev Lett 104:025701
60. Freed KF (1987) Renormalization group theory of macromolecules. Wiley, New York
61. Karplus M, McCammon J (1983) Dynamics of proteins: elements and functions. Annu Rev Biochem 53:263–300
62. Levitt M, Sanders C, Stern PS (1985) Protein normal-mode dynamics; trypsin inhibitor, crambin, ribonuclease, and lysozyme. J Mol Biol 181:423–447
63. Alexander S, Orbach RL (1982) Density of states on fractals: fractons. J Phys Lett 43:L625–L631
64. Elber R, Karplus M (1986) Low frequency modes in proteins: use of effective-medium approximation to interpret fractal dimension observed in electron-spin relaxation meaurements. Phys Rev Lett 56:394–397
65. Burioni R, Cassi D (1996) Universal properties of spectral dimension. Phys Rev Lett 76:1091–1093
66. Burioni R, Cassi D, Fontana MP, Vulpiani A (2002) Vibrational thermodynamic instability of recursive networks. Europhys Lett 58:806–810
67. Burioni R, Cassi D, Cecconi F, Vulpiani A (2004) Topological thermal instability and length of proteins. Proteins Struct Funct Bioinf 55:529–535
68. Leitner DM (2009) Frequency-resolved communication maps for proteins and other nanoscale materials. J Chem Phys 130:195101
69. Chen J, Bryngelson JD, Thirumalai D (2008) Estimations of the size of nucleation regions in globular proteins. J Phys Chem B 112:16115–16120
70. Frauenfelder H, Parak F, Young RD (1988) Conformational substates in proteins. Annu Rev Biophys Chem 17:451–479
71. Allen JP, Colvin JT, Stinson DG, Flynn CP, Stapleton HJ (1982) Protein conformation from electron spin relaxation data. Biophys J 38:299–310
72. Stapleton HJ, Allen JP, Flynn CP, Stinson DG, Kurtz SR (1980) Fractal form of proteins. Phys Rev Lett 45:1456–1459
73. Helman JS, Coniglio A, Tsallis C (1984) Fractons and the fractal structure of proteins. Phys Rev Lett 53:1195–1197
74. Herrmann HJ (1986) Comment on fractons and the fractal structure of proteins. Phys Rev Lett 56:2432
75. Stapleton HJ (1985) Comment on fractons and the fractal structure of proteins. Phys Rev Lett 54:1734
76. Liebovitch LS, Fischbary J, Koniarek JP, Todorova I, Wang M (1987) Fractal model of ion-channel kinetics. Biochim Biophys Acta 869:173–180
77. Liebovitch LS, Fischbary J, Koniarek J (1987) Ion channel kinetics: a model based on fractal scaling rather than multistate markov processes. Math Biosci 84:37–68
78. Liebovitch LS, Sullivan JM (1987) Fractal analysis of a voltage-dependent potassium channel from cultured mouse hippocampal neurons. Biophys J 52:979–988
79. Korn SJ, Horn R (1988) Statistical discrimination of fractal and Markov models of single channel gating. Biophys J 54:871–877
80. French AS, Stockbridge LL (1988) Fractal and Markov behavior in ion channel kinetics. Can J Physiol Pharm 66:967–970
81. Millhauser G, Salpeter L, Oswald RE (1988) Diffusion models of ion-channel gating and the origin of the power-law distributions from single-channel recording. Proc Natl Acad Sci USA 85:1503–1507
82. Lauger P (1988) Internal motions in proteins and gating kinetics of ion channels. Biophys J 53:877–884
83. Liebovitch LS, Toth TI (1990) Using fractals to understand the opening and closing of ion channels. Ann Biomed Eng 18:177–194

84. Liebovitch LS, Toth TI (1991) A model of ion channel kinetics using deterministic chaotic rather than stochastic processes. J Theor Biol 148:243–267
85. Lowen SB, Teich MC (1993) Fractal renewal processes. IEEE Trans Info Theory 39:1669–1671
86. Churilla AM, Gottschalke WA, Liebovitch LS, Selector LY, Todorov AT, Yeandle S (1995) Membrane potential fluctuations of human T-lymphocytes have fractal characteristics of fractional Brownian motion. Ann Biomed Eng 24:99–108
87. Lowen SB, Liebovitch LS, White JA (1999) Fractal ion-channel behavior generates fractal firing patterns in neuronal models. Phys Rev E 59:5970–5980
88. Rodriguez M, Pereda E, Gonzalez J, Abdala P, Obeso JA (2003) Neuronal activity in the substantia nigra in the anaesthetized rat has fractal characteristics. Evidence for firing-code patterns in the basal ganglia. Exp Brain Res 151:167–172
89. Kim S, Jeong J, Kim YKY, Jung SH, Lee KJ (2005) Fractal stochastic modeling of spiking activity in suprachiasmatic nucleus neurons. J Comp Neurosci 19:39–51
90. Brooks CL, Karplus M, Pettitt BM (1988) Proteins: a theoretical perspective of dynamics, structure and thermodynamics. Wiley, New York
91. Dewey TG, Spencer DB (1991) Are protein dynamics fractal? Commun Mol Cell Biophys 7:155–171
92. Bagchi B, Fleming GR (1990) Dynamics of activationless reactions in solution. J Phys Chem 94:9–20
93. Zwanzig R (1990) Rate processes with dynamical disorder. Acc Chem Res 23:148–152
94. Dewey TG, Bann JG (1992) Protein dynamics and noise. Biophys J 63:594–598
95. Ramakrishnan A, Sadana A (1999) Analysis of analyte-receptor binding kinetics for biosensor applications: an overview of the influence of the fractal dimension on the surface on the binding rate coefficient. Biotech Appl Biochem 29:45–57
96. Goychuk I, Hänggi P (2002) Ion channel gating: a first-passage time analysis of the Kramers type. Proc Natl Acad Sci USA 99:3552–3556
97. Carlini P, Bizzarri AR, Cannistraro S (2002) Temporal fluctuations in the potential energy of proteins: noise and diffusion. Phys D 165:242–250
98. Kopelmann R (1986) Rate processes on fractals: theory, simulations, and experiments. J Stat Phys 42:185–200
99. Kopelmann R (1988) Fractal reaction kinetics. Science (Washington DC) 241:1620–1625
100. Li HQ, Chen SH, Zhao HM (1990) Fractal mechanisms for the allosteric effects of proteins and enzyme. Biophys J 58:1313–1320
101. Lewis M, Rees DC (1985) Fractal surfaces of proteins. Science 230:1163–1165
102. Argyrakis P, Kopelman R (1990) Nearest-neighbor distance distribution and self-ordering in diffusion-controlled reactions. Phys Rev A 41:2114–2126
103. Turner TE, Schnell S, Burrage K (2004) Stochastic approaches for modelling in vivo reactions. Comp Biol Chem 28:165–178
104. Berry H (2002) Monte Carlo simulations of enzyme reactions in two dimensions: fractal kinetics and spatial segregation. Biophys J 83:1891–1901
105. Yuste SB, Acedo L, Lindenberg K (2004) Reaction front in an $A + B \rightarrow C$ reaction-subdiffusion process. Phys Rev E 69:036126
106. Kosmidis K, Argyrakis P, Macheras P (2003) Fractal kinetics in drug release from finite fractal matrices. J Chem Phys 119:63–73
107. Grimaa R, Schnell S (2006) A systematic investigation of the rate laws valid in intracellular environments. Biophys Chem 124:1–10
108. Shlesinger MF (1988) Fractal time in condensed matter. Annu Rev Phys Chem 39:269–290
109. Yu X, Leitner DM (2003) Anomalous diffusion of vibrational energy in proteins. J Chem Phys 119:12673–12679
110. Leitner DM (2008) Energy flow in proteins. Annu Rev Phys Chem 59:233–259
111. Li MS, Klimov DK, Thirumalai D (2004) Finite size effects on thermal denaturation of globular proteins. Phys Rev Lett 93:268107
112. Peierls RE (1934) Bemerkungüber Umwandlungstemperaturen. Helv Phys Acta 7:S81–S83
113. Reuveni S, Granek R, Klafter J (2010) Anomalies in the vibrational dynamics of proteins are a consequence of fractal-like structure. Proc Natl Acad Sci USA 107:13696–13700
114. Lois G, Blawzdziewicz J, O'Hern CS (2010) Protein folding on rugged energy landscapes: conformational diffusion on fractal networks. Phys Rev E 81:051907
115. Sangha AK, Keyes T (2009) Proteins fold by subdiffusion of the order parameter. J Phys Chem B 113:15886–15894
116. Moret MA, Santana MC, Zebende GF, Pascutti PG (2009) Self-similarity and protein compactness. Phys Rev E 80:041908
117. Morita H, Takano M (2009) Residue network in protein native structure belongs to the universality class of three dimensional critical percolation cluster. Phys Rev E79:020901
118. Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 247:536–540
119. Bashford D, Karplus M (1990) pKas of ionization groups in proteins: atomic detail from a continuum electrostatic model. Biochemistry 29:10219–10225
120. Warshel A, Papazyan A (1998) Electrostatic effects in macromolecules: fundamental concepts and practical modeling. Curr Opin Struct Biol 8:211–217
121. Zhou Z, Payne P, Vasquez M, Kuhn N, Levitt M (1996) Finite-difference solution of the Poisson–Boltzmann equation: complete elimination of self-energy. J Comput Chem 17:1344–1351
122. Lu B, Zhang D, McCammon J (2005) Computation of electrostatic forces between solvated molecules determined by the Poisson–Boltzmann equation using a boundary element method. J Chem Phys 122:214102–214108
123. Feig M, Onufriev A, Lee M, Im W, Case E, Brooks C (2004) Performance comparison of generalized Born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. J Comput Chem 25:265–284
124. Spassov V, Ladenstein R, Karshikoff (1997) A optimization of the electrostatic interactions between ionized groups and peptide dipoles in proteins. Protein Sci 6:1190–1196
125. Petrey D, Honig B (2000) Free energy determinants of tertiary structure and the evaluation of protein models. Protein Sci 9:2181–2191
126. Shoemaker K, Kim P, York E, Stewart J, Baldwin R (1987) Tests of the helix dipole model for stabilization of alpha-helices. Nature 326:563–567
127. Åqvist J, Luecke H, Quiocho F, Warshel A (1991) Dipoles localized at helix termini of proteins stabilize charges. Proc Natl Acad Sci USA 88:2026–2030
128. Wada A (1976) The alpha-helix as an electric macro-dipole. Adv Biophys 9:1–63
129. Miyazawa S, Jernigan R (1999) Evaluation of short-range interactions as secondary structure energies for protein fold and sequence recognition. Proteins 36:347–356
130. Selvaraj S, Gromiha MM (2003) Role of hydrophobic clusters and long-range contact networks in the folding of (alpha/beta)8 barrel proteins. Biophys J 84:1919–1925
131. Berman H, Henrick K, Nakamura H (2003) Announcing the worldwide. Protein Data Bank Nat Struct Biol 10:980

132. Theiler J (1987) Efficient algorithm for estimating the correlation dimension from a set of discrete points. Phys Rev A 36:4456–4462
133. Gallivan JP, Dougherty DA (1999) Cation–π interactions in structural biology. Proc Natl Acad Sci USA 96:9459–9464
134. Glockle WG, Nonnenmacher TF (1995) A fractional calculus approach to self-similar protein dynamics. Biophys J 68:46–53
135. Cserzöa M, Vicsek T (1991) Self-affine fractal analysis of protein structures. Chaos Solitons Fractals 1:431–438
136. Isvoran A (2004) Describing some properties of adenylate kinase using fractal concepts. Chaos Solitons Fractals 19:141–145
137. Mitra C, Rani M (1993) Protein sequences as random fractals. J Biosci 18:213–220
138. Banerji A, Ghosh I (2010) Mathematical criteria to observe mesoscopic emergence of protein biochemical properties. J Math Chem 49(3):643–665. doi:10.1007/s10910-010-9760-9
139. Grosberg Y, Khokhlov AR (1994) Statistical physics of macromolecules. American Institute of Physics, Woodbury
140. Røgen P, Fain B (2003) Automatic classification of protein structure by using Gauss integrals. Proc Natl Acad Sci USA 100:119–124
141. Ramnarayan K, Bohr H, Jalkanen K (2008) Classification of protein fold classes by knot theory and prediction of folds by neural networks: a combined theoretical and experimental approach. Theor Chim Acta 119:265–274
142. Kolaskar AS, Ramabrahmam V (1983) Conformational properties of pairs of amino acids. Int J Peptide Protein Res 22:83–91
143. Alexander S (1989) Vibrations of fractals and scattering of light from aerogels. Phys Rev B 40(11):7953–7965
144. Korb JP, Bryant RG (2005) Noise and functional protein dynamics. Biophys J 89:2685–2692
145. Haliloglu T, Bahar I, Erman B (1997) Gaussian dynamics of folded proteins. Phys Rev Lett 79:3090–3093
146. Aksimentiev A, Holyst R (1999) Single-chain statistics in polymer systems. Prog Polym Sci 24:1045–1093
147. Ben-Avraham D, Havlin S (2000) Diffusion and reactions in fractals and disordered systems. Cambridge University, New York
148. Arteca GA (1994) Scaling behaviour of some molecular shape descriptors of polymer chains and protein backbones. Phys Rev E 49:2417–2428
149. Arteca GA (1995) Scaling regimes of molecular size and self-entanglements in very compact proteins. Phys Rev E 51:2600–2610
150. Flory PJ (1953) Principles of polymer chemistry. Cornell University Press, New York
151. Doi M, Edwards SF (1986) The theory of polymer dynamics. Clarendon Press, Oxford
152. Rubinstein M, Colby RH (2003) Polymer physics. Oxford University Press, New York
153. Hypertext link. http://paulbourke.net/fractals/fracdim/. Accessed 16 May 2011
154. Granek R, Klafter J (2005) Fractons in proteins: can they lead to anomalously decaying time autocorrelations? Phys Rev Lett 95:098106