Physica A 388 (2009) 4600–4608

Contents lists available at ScienceDirect

Physica A

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

Fractal protein structure revisited: Topological, kinetic and thermodynamic relationships

E. Tejera [a,*], A. Machado [a], I. Rebelo [a], J. Nieto-Villar [b,c]

[a] *Faculdade de Farmácia, Departamento de Bioquímica / Instituto de Biologia Molecular y Celular (IBMC), Universidade do Porto, Portugal*

[b] *Dpto. de Química-Física, Fac. de Química, Universidad de La Habana, Cuba*

[c] *Cátedra de Sistemas Complejos "H. Poincaré", Universidad de La Habana, Cuba*

ARTICLE INFO

Article history:
Received 14 April 2009
Received in revised form 18 May 2009
Available online 17 July 2009

Keywords:
Fractal dimension
Protein length
Thermodynamic
Folding rate
Spectral dimension
Contact map

ABSTRACT

The present work explored the definitions and calculations of fractal dimensions in protein structures and the corresponding relationships with the protein class, secondary structure contents, fold type as well as kinetic and thermodynamic parameters like the folding and unfolding rate, the folding–unfolding free energy and others. The results showed a positive correlation of some fractal exponents with the kinetic and thermodynamic variables even considering the effect of the protein length. On the other hand the influences of secondary structures types, especially the turn conformation are significant as well as the fractal exponent profiles according to class and fold types.

1. Introduction

The study of fractal properties of proteins has been carried out almost from the beginnings of fractal theory development using, for this purpose, the two common forms of protein information: the amino acid sequence and the three-dimensional structure of proteins [1–3]. With renormalization group theory the fractal properties of polymers were revisited including several aspects of the secondary structure influences and structure classification [4,5,2,3]. However the application and spectrum of relationships between fractal dimensions (FD) and areas of interest like protein folding and thermodynamic properties have been little explored. In the present work, we applied several methodologies of the FD calculation to the three-dimensional structure including a modified formulation using contact maps. We showed that some of these indexes are related to the protein folding rate, protein classes, secondary structures and other topological, kinetic and thermodynamic parameters; however, these relationships will depend on the FD definition types.

2. Theoretical background

There are several approaches to the FD calculations in the three-dimensional protein structures and, in general, we can classify them into two groups depending on whether the amino acid connectivity is considered or not.

[*] Corresponding address: Faculdade de Farmácia, Departamento de Bioquímica/Instituto de Biologia Molecular y Celular (IBMC), Universidade do Porto, Portugal.
E-mail address: edutp00@yahoo.com (E. Tejera).

0378-4371/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2009.07.015

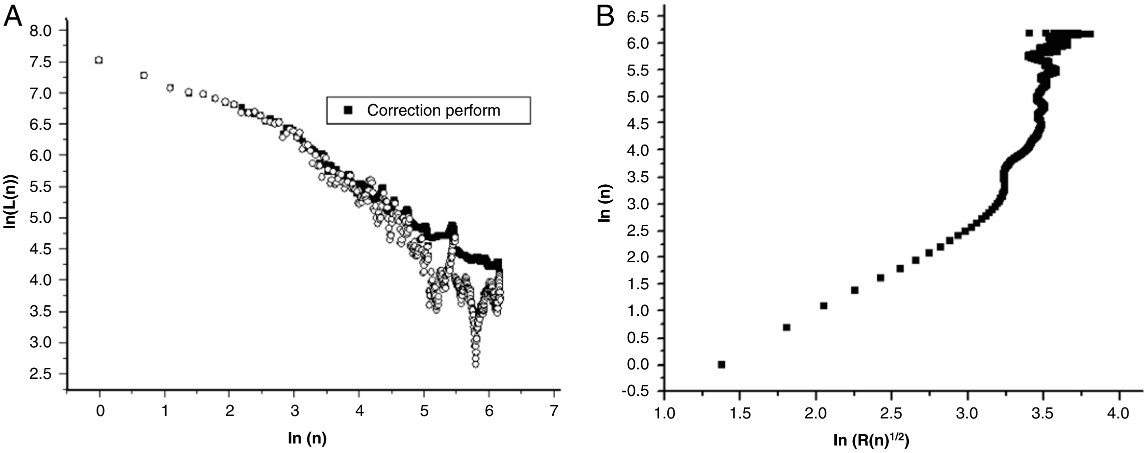

Fig. 1. (A) Scale profile for the calculation of D_1 with and without the corrections. We can note the linearity improvement with the correction mostly to high n values. (B) Scale profile for the calculation of D_2. The presented calculations were made using a carboxypeptidase protein (1AC5 PDB code) as an example.

2.1. Methods that consider the amino acid connectivity

In these kinds of methods, two similar approaches are available, and both are using the protein length as a major variable. The first method [1] defines the protein length as:

$$L(n) = L^o(n) + \frac{N - n\sigma - 1}{n\sigma} L^o(n) \quad (1)$$

where $\sigma = \text{int}\left(\frac{N}{n}\right) - 1$, N is the amino acid number and n is the length interval. The first term of this equation ($L^o(n)$) corresponds to the length of the chain for the n-integer segments while the second term is the remaining length of the segment. The fractal dimension (D_1) is then calculated by: $L(n) = B \cdot n^{1-D_1}$ in the scale-range where $\ln(L(n))$ and $\ln(n)$ are truly linearly related. This method could be improved as follows: (1) perform the $L(n)$ measure starting from different C^α instead the terminal C^α only and (2) using the actual end-to-end distance of the remaining residues. These corrections increase the quality of the regression principally for high n values [6] (Fig. 1A).

The second method [4,5,2] considers the protein length as:

$$\left\langle Rs(n)^2 \right\rangle = \frac{1}{N - n + 1} \sum_{i=1}^{N-n+1} R^2_{i,i+n} \quad (2)$$

where $R^2_{i,i+n}$, is the square of the distance between the extremes i and $i + n$, of the chain with N amino acids and n intervals like in the previous method. The fractal dimension D_2 is calculated one more time by linear regression in the scale-range appropriate to the logarithmic form of: $\left[\left\langle Rs(n)^2\right\rangle^{1/2}\right]^{D_2} = A \cdot n$.

The theoretical values of D_1 and $D_{2,}$ in the case of self-avoided walk (SAW) model, are 1.40 and 5/3 respectively [6,4,3] and the previous works have confirmed that these values are very close to those obtained in protein calculations [1,6,4,5,2,3]. However in Fig. 1A even when corrected, it was possible to note an inflexion point around $n = 15$ that was not shown in all the studied proteins. In agreement with the preceding works [6,7], we calculated D_1 in two different intervals $1 \le n \le 15$ (D_1^{1-15}) and $15 \le n \le 30$ (D_1^{15-30}). The inflexion point was observed in the second methodology (Fig. 1B) too, however in general the linearity was poor for high n values and for this reason the D_2 was calculated only in the $1 \le n \le 15$ interval. The possible differences of these scales will be discussed later.

2.2. Methods considering the spatial atomic distribution

These methods consider the proteins like clouds of points. The most common approach is the mass fractal dimension [8,9] where $\langle M(\varepsilon)\rangle \sim \varepsilon^{dm}$ and is calculated by counting the number of atoms inside the sphere of radius (ε) taking any atom of the space as centre. The average value obtained from several centres and radii is used to determine dm from the slope of $\ln(\langle M(\varepsilon)\rangle)$ with respect to $\ln(\varepsilon)$. Another methodology used in the image analysis is the classical box count fractal dimension [10] (Db) which consists of covering the three-dimensional space occupied by the object with a number of cells of ε size ($N(\varepsilon)$) and afterwards count the number of these cells that are occupied. This procedure is repeated with several cell sizes and the dimension is calculated by the slope of $\log(N(\varepsilon))$ with respect to $\log(1/\varepsilon)$ (Fig. 2B).

We exploited another way to do the calculation of the FD using the contact map and the contact number ($Nc(\varepsilon)$) definition as the starting point. The contact map matrix ($Mc(\varepsilon)$) is defined as:

$$Mc_{i,j} = \begin{cases} 1 & \text{if } d_{i,j} \leq \varepsilon \\ 0 & \text{if } d_{i,j} > \varepsilon \text{ or } i = j \end{cases} \tag{3}$$

where ε represents the cut-off distance and $d_{i,j}$ is the special distance between the residues i and j. The contact number by residues $N(\varepsilon)_i$ is then:

$$Nc(\varepsilon)_i = \sum_{j}^{N} Mc_{i,j} = \sum_{j}^{N} \theta\left(\varepsilon - \left\|d_{i,j}\right\|\right) \tag{4}$$

where N is the amino acids number and θ is the Heaviside function. The total number of contacts in the protein $N(\varepsilon)$ is:

$$Nc(\varepsilon) = \sum_{i}^{N} Nc\,(\varepsilon)_i = \sum_{i}^{N} \sum_{j \neq i}^{N} Mc_{i,j} = 2 \sum_{i<j}^{N} \theta\left(\varepsilon - \left\|d_{i,j}\right\|\right). \tag{5}$$

The last term considers the symmetrical properties of the contact matrix. The relationship between Nc, ε and FD could be easily deduced using the definition of correlation dimension of Grassberger and Procaccia [11] used in images and temporal series analysis: $C(\varepsilon) \sim \varepsilon^{Dc}$. The correlation function is defined as:

$$C\,(\varepsilon) = \frac{2}{N(N-1)} \sum_{i<j}^{N} \theta\left(\varepsilon - \left\|r_{ij}\right\|\right). \tag{6}$$

where N in this equation corresponds to the number of points and $r_{i,j}$ is the distance between points i and j. It is easy to note that in our case N is equal to the residue numbers and $r_{i,j} = d_{i,j}$. Therefore by substitution of Eq. (4) in Eq. (5):

$$C\,(\varepsilon) = \frac{1}{N(N-1)} Nc\,(\varepsilon). \tag{7}$$

Note that $N(N-1)$ is a normalization term, therefore $C(\varepsilon)$ could be considered as the probability that any two residues are in contact at a cut-off distance ε. As $C(\varepsilon) \sim \varepsilon^{Dc}$ then $Nc(\varepsilon) \sim \varepsilon^{Dc}$ and the fractal dimension (Dc) is calculated by the slope of $\ln(C(\varepsilon))$ vs $\ln(\varepsilon)$ (Fig. 2A).

2.3. Relationship between dm and Dc

As we saw before, $\langle M(\varepsilon)\rangle \sim \varepsilon^{dm}$, however $M(\varepsilon)$ to the centre I, which could be calculated as:

$$M(\varepsilon)_i = \sum_{j}^{N} \theta\left(\varepsilon - \left\|d_{i,j}\right\|\right). \tag{8}$$

That is exactly the number of atoms inside the radio ε and centre i. If we make the sum over all the centres and using Eq. (6), the average mass for a given radio is:

$$\langle M(\varepsilon)\rangle = \frac{1}{N} \sum_{i}^{N} \sum_{j \neq i}^{N} \theta\left(\varepsilon - \left\|d_{i,j}\right\|\right) = \frac{1}{N} \sum_{i}^{N} \sum_{j \neq i}^{N} Mc_{i,j} = \frac{1}{N} Nc(\varepsilon) = (N-1)C(\varepsilon). \tag{9}$$

From Eq. (8) we note that if $Nc(\varepsilon) \sim \varepsilon^{Dc} \sim \varepsilon^{dm}$, then the slope of $\ln(\langle M(\varepsilon)\rangle)$ vs $\ln(\varepsilon)$ is equal to $dm = Dc$. It is evident that the last one is true for very big structural systems where the frontier or limited size effect is not present. However the residues located in the protein surface could affect the $dm = Dc$ equality, in fact, the mass fractal dimension is often calculated in residues close to the centre of mass [8,9]. The convenience of defining the mass fractal dimension as a function of the contact map bring firstly a comfortable formulation and secondly the contact map has an intuitive physical meaning and is used in several formulation related to folding and structure prediction as well as graph theory where an arsenal of theoretical and mathematical tools are available.

2.4. Spectral dimension

The spectral dimension could be classified in the second group of methods because the residue connectivity is not implicit in the formulation. The calculation of the spectral dimension is performed on the basis of Laplacian or Kirchhoff matrices (Γ):

$$\Gamma = -Mc + D \tag{10}$$

where D represents the matrix degree. The spectral dimension is calculated by the relation $G(\omega) \sim \omega^{Ds-1}$ where $G(\omega)$ denotes the density of low frequency modes calculated from the eigenvalues of Γ. In the present work, two cut-off values 6 and 7 Å were used to calculate Ds^6 and Ds^7 respectively using the first 50 modes [9,12].

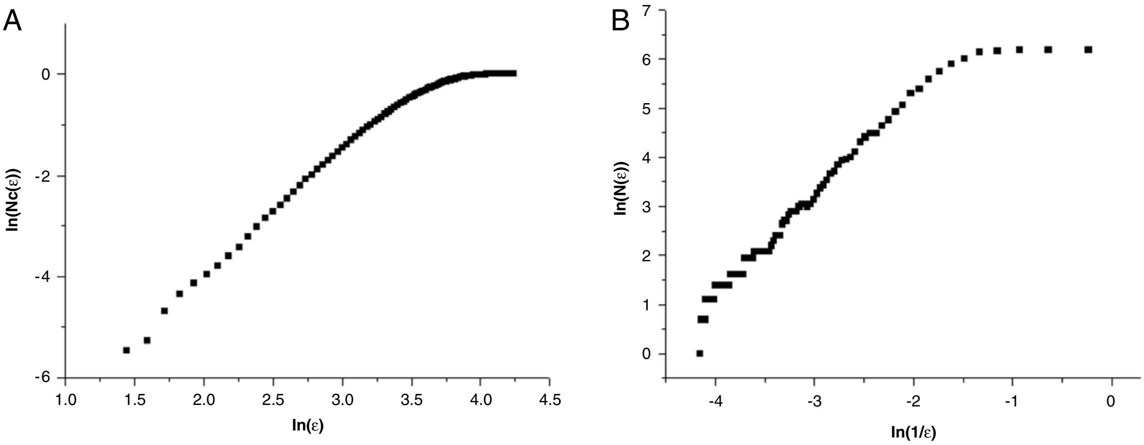

Fig. 2. (A) Scale profile for the calculation of *Dc*. We can note a wide linear interval. (B) Scale profile for the calculation of *Db*. The presented calculation was made using a carboxypeptidase protein (1AC5 PDB code) as example.

2.5. Contact order

Besides the fractal dimensions we calculate the contact order (CO) defined as [13]:

$$CO = \frac{1}{N \cdot Nc} \sum_{ij}^{N} \Delta L_{ij}. \tag{11}$$

where N was the residue number, Nc was the contact number for a predefined cutoff of 6 Å and ΔL_{ij} was the number of residues between i and j that are in contact.

3. Protein group selected and secondary structure calculation

We selected a total of 870 proteins from the Protein Data Bank [14] with X-Ray diffraction as the structure elucidation method, a resolution of less than 2.5 Å and composed of only one chain. The protein length interval of selected proteins is 198–937 (min–max) residues with an average of 314 ± 99 residues. The class and fold assignment were according to the SCOP database [15]. The calculation of fractal dimensions, contact order, and amino acids and secondary structure percent, were done with Pascal home made software. The complete set of proteins is available in the Supplementary Materials I (see Appendix).

3.1. Thermodynamic and kinetic data

For the study of the relationships between fractal exponents, folding kinetic and thermodynamic parameters, we extracted a protein set from several articles. The complete list is presented in the Supplementary Materials II (see Appendix).

3.2. Secondary structure content

The secondary structure assignment was performed using the DSSP software [16]. We considered the residue number in: α-helix (H), the extended strand, participates in β-ladder (E), bends (S), H-bonded turns (T), 3_{10}-helixes (G) and the residues in isolated β-bridges.

4. Results and discussion

The relationships between the fractal dimension values calculated using the different methods as well as the values obtained for the complete set of proteins, revealed two groups with different properties: a first group Dc, Db, dm were highly inter-correlated with increased values and the other one D_1^{1-15} and D_2 with similar inter-correlations but lower exponent values (Tables 1 and 2). The reasons of these differences were associated with the consideration or not of the amino acids' connectivity [5,2]. In the first group the scaling exponent was referred to the geometrical radius whereas the considerations of connectivity resulted in scaling exponent with respect to the end-to-end distance.

The Ds values are lower than Dc, Db or dm however the correlation within this group is higher with respect to the second one (Table 2). The values obtained for D_1^{1-15} and D_2 are very close to the SAW model as has been confirmed by

Table 1
Global median values of fractal dimensions calculated in the complete protein set.

Dc	2.58 (2.54–2.62)
Db	2.05 (1.98–2.12)
D_1^{1-15}	1.38 (1.36–1.4)
D_2^{1-15}	1.54 (1.5–1.58)
D_1^{15-30}	1.97 (1.8–2.12)
Ds^7	1.86 (1.82–1.9)
Ds^6	1.72 (1.67–1.76)
dm	2.85 (2.8–2.9)

The values between (. . .) represent the 25 and 75 quartiles interval.

Table 2
Correlation values between the fractal dimensions calculated by several methods.

	Dc	*CO*	*Db*	D_1^{1-15}	D_2	D_1^{15-30}	Ds^7	Ds^6	*Dm*
Dc	**1.00**		0.46				0.61	0.61	0.70
CO		**1.00**	0.11	0.08	−0.08	0.18	0.32	0.31	
Db	0.46	0.11	**1.00**				0.44	0.40	0.38
D_1^{1-15}		0.08		**1.00**	0.76	0.10	0.21	0.14	
D_2		−0.08		0.76	**1.00**		0.14	0.07	
D_1^{15-30}		0.18		0.10		**1.00**	0.18	0.14	−0.07
Ds^7	0.61	0.32	0.44	0.21		0.18	**1.00**	0.87	0.42
Ds^6	0.61	0.31	0.40	0.14		0.14	0.87	**1.00**	0.41

Blank space cells represent insignificant correlation ($p > 0.05$).

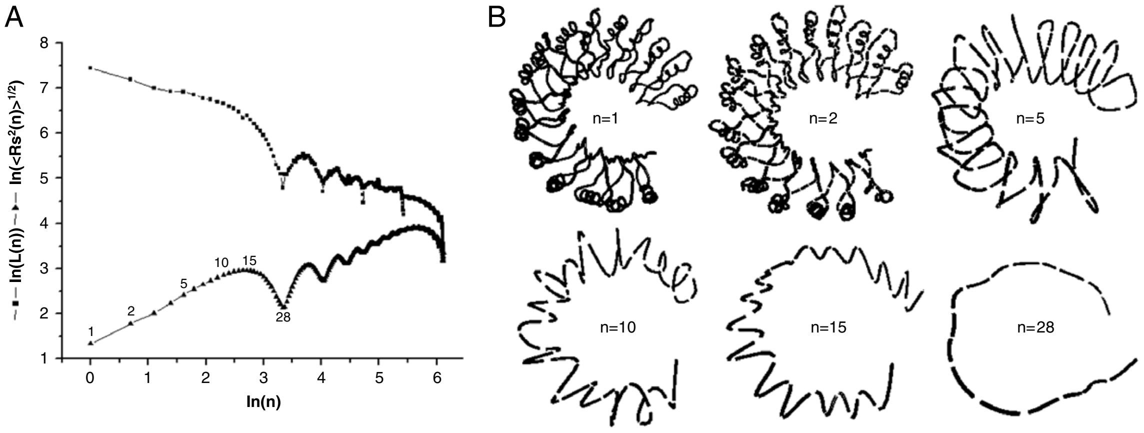

Fig. 3. (A) Representation of Eqs. (1) and (2) (exchanges axis). (B) Approximated representations of the protein structure using different scales. Both equations follow the same pattern corresponding to different representations of the protein structure. The represented calculation was made using the porcine ribonuclease inhibitor protein (PDB code: 2BNH) as an example.

other authors [1,6,4,5,2,3,7–9]. However the D_1^{15-30} values are higher that of the SAW model in fact, this value is higher than the corresponding ones of unrestricted random walk (URW) model (1.5 approximately). This could lead to the conclusion that the proteins are more compact as we would expect from a URW chain where no interaction forces are present, however it is contrary to all the other fractal exponents where they remain clearly as an intermediary state between poor (or good solvent like) and high (bad solvent) compactness.

The D_1^{15-30} index is poorly correlated with all the other indexes and this could be caused by two possible factors: algorithmic error or different scale behaviour. The inflexion point noted in the D_2 calculation (Fig. 3) is not shown in all the proteins and remains with the algorithm correction. On the other hand it was not present in the methods that did not consider the amino acid connectivity.

If the scale length (n) increases, the secondary structure geometry influence decreases, the helixes and turns, for example, disappear, emerging an overall geometry that could has a different self-similarity rule (Fig. 3B) and therefore, could be regulated by another scale exponents. Over the $n = 28$ (in the Fig. 3 example) the geometry (almost a curve) will change increasing the roughness similar to $n = 10$. This effect begins with small scale values when the limited size effect ($n \sim N$) is not predominant. Multifractal patterns have been found in protein sequences [17,18] however in protein structure it is almost unexplored.

Table 3
Median values of the studied variables according to the class groups.

	α	β	α/β	$\alpha+\beta$
Dc[c]	2.55 (2.5–2.62)	2.58 (2.52–2.62)	2.59 (2.55–2.62)	2.57 (2.53–2.6)
Db[c]	2.02 (1.95–2.11)	2.06 (1.97–2.13)	2.06 (1.99–2.13)	2.02 (1.96–2.11)
D_1^{1-15}[a]	1.37 (1.34–1.4)	1.40 (1.37–1.43)	1.38 (1.36–1.4)	1.39 (1.36–1.41)
D_2[c]	1.54 (1.48–1.58)	1.52 (1.47–1.57)	1.54 (1.5–1.58)	1.55 (1.49–1.6)
D_1^{15-30}[b]	1.72 (1.61–1.83)	2.06 (1.91–2.26)	1.98 (1.85–2.13)	2.00 (1.83–2.11)
Ds^7[b]	1.80 (1.75–1.88)	1.90 (1.83–1.95)	1.86 (1.82–1.9)	1.85 (1.82–1.89)
Ds^6[a]	1.67 (1.6–1.73)	1.74 (1.7–1.8)	1.72 (1.68–1.76)	1.70 (1.66–1.75)
dm[c]	2.83 (2.71–2.91)	2.83 (2.78–2.87)	2.87 (2.82–2.91)	2.84 (2.79–2.88)

The values between (. . .) represent the 25th and 75th quartile intervals.
[a] Significant differences ($p < 0.01$) between all groups.
[b] Significant differences ($p < 0.01$) between all groups except between the groups: (α/β) and ($\alpha + \beta$).
[c] See the discussion.

According to Eq. (9) we must expect that $Dc \cong dm$, however $Dc < dm$ (Table 1). This is a consequence of the different residue distributions in the core and surface. The surface residues have a less compact environment and consequently tend to decrease the fractal dimension contrary to core residues.

4.1. Class and secondary structure relationships

As was referred to and according to the previous works, the D_1^{1-15} values must depend on the secondary structure type. Ideally secondary structure D_1^{1-15} values are: 1.44 (0.13), 1.09 (0.06), 1.06 (0.04) and 1.07 (0.05) for the α-helix, parallel β-sheet, anti-parallel and twisted β-sheet respectively [6,2,7]. In the α-helix and reverse turn geometry the distance separation between not neighboring residues is less, increasing the fractal dimension contrary to the ordered β-sheet.

In the previous works [6] with a dataset of 90 proteins, the D_1^{1-15} values for β class were relatively larger (1.33) with respect to the ideal value while α class was around 1.41. In our results, only D_1^{1-15}, D_1^{15-30}, Ds^6 and Ds^7 are significantly different between α and β classes with a relative higher values of β with respect to α class (Table 3). The other exponents are only different with respect to α and β mixed classes. However, the D_2 values show significant differences between (α/β), $(\alpha + \beta)$ and β classes but the influence of α group cannot be discriminated. On the other hand the Dc and dm inter-group differences are only significant with respect to (α/β), corresponding to the higher dimensions, that could indicate a predominant compact core in this protein family.

The influence of turns on the fractal exponent is positive and relatively strong however, the influence of α and β percentage are in disagreement with ideal values and the previous works. One of the possible reasons of these deviations is associated with the turn percent. We can note that higher dimension values correspond to proteins with high content of turns in α-class as well as β-class. In general however, the frequent presence of β structures (and low α-helix) lead to elevated compositions of turns (>20%) that increase considerably the dimension values (Fig. 4). This effect could explain the increased values of β-class.

Even when the turn influence could explain the increment of fractal exponents in the β-class, a global explanation including the α-class behavior is more difficult because in fact, the exponent variations are fold type dependent. The fractal exponents only do not depend of the composition but they are also influenced by the location and deformation of the secondary structures. In out dataset, there are 182 different fold types (following the SCOP classification) with a very wide secondary structure distribution and dimensions (Table 4).

Proteins with approximately the same secondary structure percentage could have different fractal exponents and vice versa. The annexins (47 873) and cytochrome c (46 625) fold types have similar values of Dc and D_1^{1-15} (Table 4), however the secondary structure percentage is different. On the other hand, with respect to α/α toroid (48 207) a contrary relationship is noted. The explanation of these differences is not simple, even when a wide distribution of the α-helixes around the protein core could explain the fractal exponent increment; the same thing could be said concerning to the α-helixes on the surface. Similar pattern could be noted comparing the 7-bladed beta-propeller (50 964) and Immunoglobulin-like beta-sandwich (48 725) fold types.

4.2. Kinetic and thermodynamic relationships

The influence of the native structure topology on the folding rate is a well-known phenomenon. The two-state folding proteins show a correlation with CO and several others topological parameters related to the protein length and the secondary structure [19–21]. On the other hand in the three or multiple-state folding proteins this correlation with CO is poor and the folding rate ($\ln k_f$) is mostly related to the protein length [22–24] (N) as a general power law $\ln k_f \sim N^{\alpha}$. Others topological descriptors based on graph analysis had been related to the folding–unfolding free energy or the unfolding rate [25], however, the relationship between the fractals exponent an all theses kinetic and thermodynamic variables are little explored.

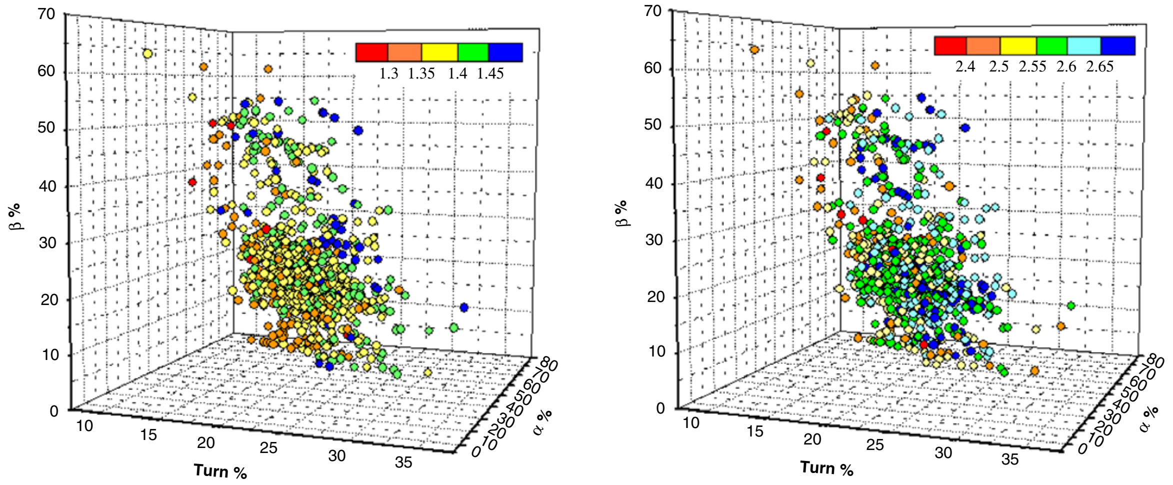

Fig. 4. Distribution of fractal exponent values considering the alpha, beta and turn percent. (Left) The colour scale is related to D_1^{1-15}. (Right) the colour scale is related to *Dc*. We can note that the variation of fractal exponent with respect to secondary structure content is complex and sensitive to the turns-percent. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Results for some of the fold types present in the dataset.

Fold code	Class	Dc	Db	D_1^{1-15}	D_2	Ds^6	dm	α-%	β-%	Turn-%
46625	α	2.55	2.07	1.38	1.54	1.72	2.83	38.71	8.53	23.54
47873		2.54	1.96	1.38	1.56	1.66	2.81	71.66	0.00	13.86
48112		2.63	2.10	1.39	1.55	1.74	2.89	46.57	6.59	24.10
48207		2.62	2.04	1.40	1.62	1.73	2.89	54.92	7.32	20.47
48725	β	2.46	1.81	1.38	1.42	1.68	2.51	2.18	48.84	21.98
51181		2.56	1.94	1.36	1.49	1.69	2.81	28.36	27.95	20.47
50629		2.59	2.08	1.44	1.59	1.78	2.83	12.85	46.89	22.17
50938		2.64	2.05	1.42	1.57	1.77	2.97	7.87	46.29	22.74
50964		2.64	2.00	1.44	1.58	1.70	3.01	2.95	51.90	26.96
50933		2.65	2.07	1.42	1.52	1.71	2.96	3.15	49.00	27.14
52046	$\alpha\beta$	2.51	2.02	1.43	1.61	1.88	2.56	14.33	21.14	28.71
55619		2.52	2.03	1.34	1.47	1.66	2.73	31.70	33.25	18.08
52373		2.56	2.06	1.35	1.49	1.66	2.86	55.24	10.09	16.56
52539		2.56	2.00	1.36	1.46	1.68	2.84	41.30	23.03	17.17
51350		2.61	2.07	1.37	1.52	1.73	2.89	43.06	16.83	20.90
53849		2.61	2.05	1.40	1.59	1.76	2.89	40.40	18.90	22.39
52732		2.68	2.05	1.38	1.52	1.77	2.88	51.55	11.05	17.97
53162		2.66	2.06	1.37	1.49	1.76	2.93	37.58	18.29	24.76

The fold and class type classifications are according to the SCOP 1.73 notation. The values are the average of those proteins corresponding to the same fold type and with a residue number close to 300 (298–340).

Table 5
Correlation coefficient between fractal exponents and several folding kinetics and thermodynamics parameters.

	$\ln k_F$	$\ln k_U$	ΔG_{N-U}^{eq}	m_F	m_U	m_{eq}	β_{TS}
Db						−0.51	
CO	−0.29						
Dc	−0.52	−0.61	0.49	−0.38	0.53	−0.54	−0.39
D_2	0.27	0.45					0.40
D_1^{1-15}						0.47	0.28
dm	−0.46	−0.58	0.62	−0.33	0.51	−0.48	−0.43
Ds^7				0.29	−0.36		

Blank space cells represent insignificant correlations. $\ln k_F$, $\ln k_U$: Natural logarithm of protein folding and unfolding rate respectively. ΔG_{N-U}^{eq}: The free energy of unfolding in water. m_F, m_U: The dependence of natural logarithms of folding and unfolding rates respectively on denaturant concentration. m_{eq}: The dependence of the free energy of unfolding on the denaturant concentration. β_{TS}: Position of transition state in the reaction coordinates.

The *Dc* and *dm* show a wide correlation with several folding parameters contrary to the fractal exponent based on protein connectivity. In these cases where a correlation is present simultaneously with both kinds of fractal exponent, a contrary effect is noted (Table 5). Once more this is a consequence of the different physical meaning involved in the calculations.

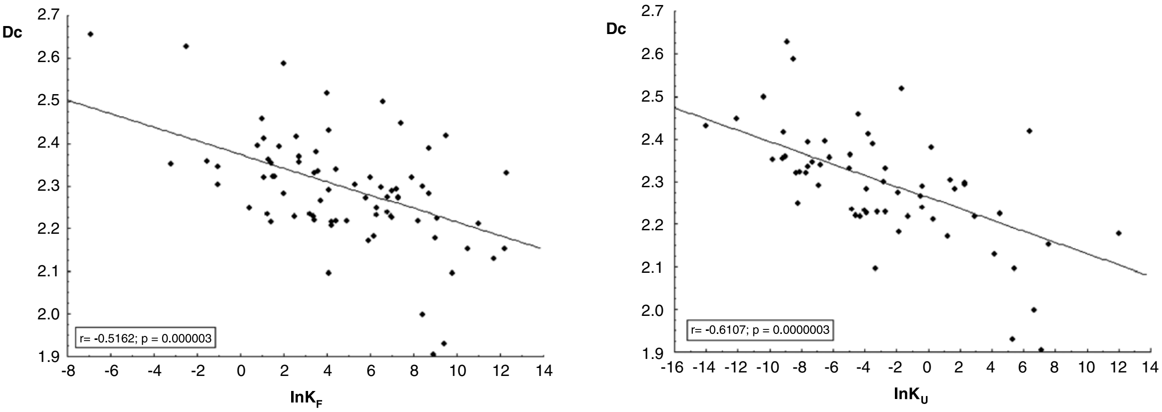

Fig. 5. Fractal dimension variation (Dc) with respect to: (Left) folding rate ($\ln k_F$) and (Right) Unfolding rate (lnk_U)..

The increment in the fractal exponent is associated with a reduction in the folding and unfolding rate (Table 5, Fig. 5). The data presented in Fig. 5 as well as the calculated correlation coefficients ($r = -0.52$ and $r = -0.61$ for $\ln k_F$ and $\ln k_U$ respectively) do not exclude the effect of the protein length however performing a partial correlation using the protein length as a control variable the correlation is reduced but remain significant ($r = -0.43$ and $r = -0.58$ for $\ln k_F$ and $\ln k_U$ respectively). The correlation differences between Dc and dm are small except for the ΔG^{eq}_{N-U} as a consequence of the influence of surface residues. The reduction of correlation by surface residues inclusion could be explained considering that the variation of compactness associated with the residue inclusion is not homogeneous; this means that the surface residues inclusion has not the same effect in all the proteins. The high correlation of dm suggests that the core topology has a major effect on the folding thermodynamic and kinetic properties.

The denaturant concentration increment is associated with a decrement in the folding rate ($m_F < 0$) and this influence is deeper in low density core proteins. The denaturant can deform the native state easily in those proteins where the molecular accessibility is higher and therefore less compact residues distribution. The influence of secondary structure (principally turns and α-helixes) is important to the final value of fractal exponent and its in agreement with the previous works that showed that the folding rate and transition state position could be predicted considering the number of residues in α-helixes structure [26,27]. This alpha influence is not a hierarchy of secondary structure formation; many pathways could be followed in the energy landscape altering the secondary order formation because many of these local structures are present in the unfolded and transition states.

The fractal exponent similarities in several folds types suggest that the local environment and interactions of different secondary structures need to be considered as well as the secondary structure content in the relationships between topology and folding rate, principally because, many of these aspects (principally in the core) are similar in the transition state structure.

5. Conclusions

In the present work we explored several definitions and calculations of fractal dimension in protein structures as well as their relationships with the protein classes, fold types and kinetic and thermodynamic parameters.

Most of the fractal exponents are significantly different for α- and β-class proteins; however those exponents calculated considering amino acid connectivity seems to be more accurate for α- and β-class differentiation. The presence of turns has a considerable influence on the fractal exponent calculation; tending to increase the values in all the classes. The fractal exponent values are dependent on the fold type and therefore on the location and connectivity of the secondary structures.

We have shown correlations of several fractal exponents with the folding/unfolding rate, folding/unfolding free energy, and position of transition state in the reaction coordinates as well as other thermodynamic and kinetic variables that help to integrate compactness properties associated with the fractal exponents.

On the other hand, several aspects related to scales anomalies were explored suggesting a possible structural multifractal behaviour at least in some proteins and mainly presented if the residue connectivity is considered, this means, if the measurement is performed on base of the end-to-end protein length.

Acknowledgment

This study was supported by "Fundação para a Ciência e a Tecnologia" (FCT), grant: SFRH/BD/25167/2005.

Appendix. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.physa.2009.07.015.

References

[1] Y. Isogai, T. Itoh, Fractal analysis of tertiary structure of protein molecule, J. Phys. Soc. Japan 53 (1984) 2162.

[2] Gerald C. Wagner, J. Trevor Colvin, James P. Allen, Harvey J. Stapleton, Fractal models of protein structure, dynamics, and magnetic relaxation, J. Am. Chem. Soc. 107 (1985) 20.

[3] S. Havlin, D. Ben-Avraham, New approach to self-avoiding walks as a critical phenomenon, J. Phys. A: Math. Gen. 15 (1982) L321–L328.

[4] S. Havlin, D. Ben-Avraham, Fractal dimensionality of polymer chains, J. Phys. A: Math. Gen. 15 (1982) L311–L316.

[5] J.T. Colvin, H.J. Stapleton, Fractal and spectral dimensions of biopolymer chains: Solvent studies of electron spin relaxation rates in myoglobin azide, J. Chem. Phys. 82 (1985) 10.

[6] Yi Xiao, Comment on fractal study of tertiary structure of proteins, Phys. Rev. E 46 (1994) 6.

[7] C.X. Wang, Y.Y. Shi, F.H. Huang, Fractal study of tertiary structure of proteins, Phys. Rev. A 41 (1990) 12.

[8] Matthew B. Enright, David M. Leitner, Mass fractal dimension and the compactness of proteins, Phys. Rev. E 71 (2005) 011912.

[9] Sh. Reuveni, R. Granek, J. Klafter, Proteins: Coexistence of stability and flexibility, Phys. Rev. Lett. 100 (2008) 208101.

[10] B.B. Mandelbrot, The Fractal Geometry of Nature, Freeman, New York, 1982.

[11] P. Grassberger, I. Procaccia, Characterization of strange attractors, Phys. Rev. Lett. 50 (5) (1983) 346–349.

[12] R. Burioni, D. Cassi, F. Cecconi, A. Vulpiani, Topological thermal instability and length of proteins, Proteins: Structure, Function, and Bioinform. 55 (2004) 529–535.

[13] K.W. Plaxco, K.T. Simons, D. Baker, Contact order, transition state placement and the refolding rates of single domain proteins, J. Mol. Biol. 277 (1998) 985–994.

[14] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Research 28 (2000) 235–242. http://www.pdb.org.

[15] A.G. Murzin, S.E. Brenner, T. Hubbard, C. Chothia, SCOP: A structural classification of proteins database for the investigation of sequences and structures, J. Mol. Biol. 247 (1995) 536–540. http://scop.mrc-lmb.cam.ac.uk/scop/.

[16] W. Kabsch, Ch. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (12) (1983) 2577–2637.

[17] T.G Dewey, B.J. Strait, Multifractals, encoded walks and the ergodicity of protein sequences, Pac. Symp. Biocomput. (1996) 216–229.

[18] Zu-Guo Yu, Vo Anha, Ka-Sing Lau, Chaos game representation of protein sequences based on the detailed HP model and their multifractal and correlation analyses, J. Theoret. Biol. 226 (3) (2004) 341–348.

[19] M.M. Gromiha, S. Selvaraj, Inter-residue interactions in protein folding and stability, Progr. Biophys. Mol. Biol. 86 (2004) 235–277.

[20] Dmitry N. Ivankov, Sergiy O. Garbuzynskiy, Eric Alm, Kevin W. Plaxco, David Baker, Alexei V. Finkelstein, Contact order revisited: Influence of protein size on the folding rate, Protein Sci. 12 (2003) 2057–2062.

[21] B. Nolting, W. Schalike, P. Hampel, F. Grundig, S. Gantert, N. Sips, W. Bandlow, P.X. Qi, Structural determinants of the rate of protein folding, J. Theoret. Biol. 223 (2003) 299–307.

[22] N. Koga, Shoji Takada, Roles of native topology and chain-length scaling in protein folding: A simulation study with a go-like model, J. Mol. Biol 313 (2001) 171–180.

[23] O.V. Galzitskaya, S.O. Garbuzynskiy, D.N. Ivankov, A.V. Finkelstein, Chain length is the main determinant of the folding rate for proteins with three-state folding kinetics, Proteins: Structure, Function, and Genetics 51 (2003) 162–166.

[24] David De Sancho, Urmi Doshi, Victor Muñoz, Protein folding rates and stability: How much is there beyond size? J. Am. Chem. Soc. 131 (6) (2009) 2074–2075.

[25] Jaewoon Jung, Jooyoung Lee, Hie-Tae Moon, Topological determinants of protein unfolding rates, Proteins: Structure, Function, and Bioinform. 58 (2005) 389–395.

[26] Dmitry N. Ivankov, Alexei V. Finkelstein, Prediction of protein folding rates from the amino acid sequence-predicted secondary structure, PNAS 101 (24) (2004) 8942–8944.

[27] Igor B. Kuznetsov, Shalom Rackovsky, Class-specific correlations between protein folding rate, structure-derived, and sequence-derived descriptors, Proteins: Structure, Function, and Bioinform. 54 (2004) 333–341.