

## MODELO DE CRITICALIDADE AUTO-ORGANIZADA PARA O DOBRAMENTO DE PROTEÍNAS COM REPETIÇÃO EM TANDEM

Fernando S. Silva<sup>1</sup>, Marcelo A. Moret<sup>2</sup> e Marcelo C. Santana<sup>3</sup>

<sup>1</sup>Departamento de Químicas e Exatas - UESB e Faculdade de Tecnologia SENAI CIMATEC, E-mail: fermathem@yahoo.com.br;

<sup>2</sup>Faculdade de Tecnologia SENAI CIMATEC, E-mail: mamoret@gmail.com;

<sup>3</sup>Universidade Salvador - UNIFACS, E-mail: fermath@gmail.com.

### RESUMO

*As proteínas são biopolímeros essenciais a vida, constituídas por uma longa sequência de aminoácidos. A presença de repetições em tandem de aminoácidos, de 20 a 50 aminoácidos, tem despertado grande interesse em diversas áreas. As proteínas de repetição em tandem possuem como vantagem os percursos dobráveis que são passíveis de serem planejadas. Este artigo propõe a aplicação da  $q$ -Gaussiana para justificar o comportamento fractal de 1213 proteínas de repetição em tandem. Nossos resultados mostram que a criticalidade auto-organizada é uma explicação para enovelamento de proteínas.*

**Palavras-Chave:** *Enovelamento; Criticalidade; Fractal;*

### ABSTRACT

*Proteins are biopolymers essential to life, constituted by a long amino acid sequence. The presence of tandem repeats of amino acids, 20-50 amino acids, has aroused great interest in several areas. The tandem repeat proteins have the advantage of the folding pathways that are likely to be planned. This article proposes the application of  $q$ -Gaussian to justify the fractal behavior 1213 repeat proteins in tandem. Our results show that the self-organized criticality is an explanation for protein folding.*

**Keywords:** *Folding; Criticality; Fractal;*

## 1. INTRODUÇÃO

As proteínas são indispensáveis para a vida, direta ou indiretamente, através da realização de várias funções em todos os processos biológicos mantenedores da vida. A sequência dos aminoácidos de uma proteína adota única configuração geométrica tridimensional: o estado nativo ou enovelado, e esta estrutura tridimensional determina a sua função biológica. Dessa forma, trata-se de um fenômeno molecular complexo que depende da ação cooperativa de muitas interações não ligadas [1].

A estrutura das proteínas dobradas foi observada mediante raios-X e ressonância magnética nuclear, descobrindo-se diferentes formas irregulares extremamente complexas [2,3], em que a Geometria Euclidiana ou a Diferencial não fornecem meios para extrair uma medida, conceito básico originário da Geometria, sobre as formas irregulares. Assim, foi necessário escolher uma geometria mais adequada para a análise de conformações das proteínas.

O conceito de Geometria Fractal foi aplicado com sucesso a um número significativo de propriedades das proteínas [4], cujo objetivo lida com as formas irregulares, por exemplo, a forma de nuvem no céu ou da linha de costa no mapa etc., fornecendo meios para extrair uma regra ou uma regularidade escondida nas formas irregulares.

Um modelo de campo médio foi proposto para proteínas globulares [5] e também para proteínas com repetição em tandem [6]. Nesses trabalhos, a energia é a soma de uma energia elástica mais uma energia repulsiva, encontrando uma lei de potência  $M \propto R_g^\delta$ ,  $\delta$  expoente de escalonamento com  $\delta = 2.15$  para proteínas com repetição em tandem. Dessa forma, as proteínas de repetição em tandem apresentam um comportamento não euclidiano evidenciando uma geometria entre um arame irreversivelmente amassado no plano e a de esferas aleatórias no limiar de percolação [6, 5].

O principal motivo que leva uma proteína ao enovelamento são as forças hidrofóbicas [7]. As macromoléculas estão sob ação de forças de dobras e de repulsão, devido à sua constituição eletrônica. Segundo as leis da termodinâmica um sistema de partículas tende a seguir os caminhos de menor consumo de energia até que encontre um ponto de equilíbrio e ali permanecerá enquanto não for perturbado. O entendimento dos mecanismos de auto-organização foi proposto por Bak [8], ilustrando com o modelo da pilha de areia o conceito da criticalidade auto-organizada (SOC) em um sistema complexo.

Neste artigo apresentamos uma distribuição cauda longa para representar um sistema complexo como os fractais. É sabido que as distribuições gaussianas, em geral, não se ajustam bem a objetos fractais [8]. Uma solução inteligente e bastante utilizada são as distribuições  $q$ -Gaussianas, pois o índice  $q$  está diretamente relacionado com a dimensão fractal do movimento aleatório [9]. Dentro desse

contexto, o objetivo do estudo é investigar as características geométricas das 1213 cadeias de proteínas comparando volume e massa de cada cadeia de proteína.

## 2. METODOLOGIA

Neste artigo foram selecionados, para investigar as características geométricas, 1213 proteínas de seis classe de repetição em tandem [6,11]: Ankyrin, Armadillo Heat,  $\beta$ -propeller,  $\beta$ -trefoil, Tetratricopeptide e Leucine Rich, coletadas no banco de dados virtual de proteínas “*Brookhaven Protein Data Bank*” (PDB) e sugeridas por Andrade e colaboradores [13].

A estratégia foi a de encontrar a densidade média de compactação massa-volume, com volume esférico aproximado pelo raio de giro  $R_g$  ( $V \approx \frac{4}{3}\pi R_g^3$ ), para analisarmos a criticalidade auto-organizada. Para tanto usaremos a função q-Gaussiana e a entropia de Tsallis.

### Função q-GAUSSIANA e Entropia de Tsallis

Neste tópico apresentamos uma breve introdução à família de funções exponenciais generalizadas. Borges, em 1998, propõe a generalização dessas funções com base na estatística de Tsallis [14]. Nesse formalismo, a função exponencial foi generalizada para uma função denominada  $q$ -exponencial dada por

$$\exp_q(x) \equiv [1 + (1 - q)x]^{\frac{1}{1-q}}, \quad (1)$$

se  $[1 + (1 - q)x] > 0$  e zero caso contrário. O parâmetro real  $q$  é conhecido como índice entrópico associado à entropia de Tsallis [9,10] que postulou uma nova entropia, partindo da generalização da entropia de Boltzmann-Gibbs-Shannon (BGS) dada por

$$S_q = k \frac{1 - \sum_{i=1}^W (p_i)^q}{q - 1} \quad (2)$$

onde  $k$  é uma constante positiva, um parâmetro real e  $\{p_i\}$  é um conjunto de probabilidades associadas a  $W$  microestados do sistema, isto é,  $\sum_{i=1}^W p_i = 1$ .

É precisamente a linguagem formal da Matemática utilizada nas equações (1) e (2) que permite uma simplicidade atraente no caso limite,  $\lim_{q \rightarrow 1} S_q = -k \sum_{i=1}^W p_i \ln p_i$  e

$\lim_{q \rightarrow 1} \exp_q(x) = \exp_1(x) = e^x$  retomam a entropia de BGS e a função exponencial padrão.

Uma propriedade importante de  $S_q$  surge quando um sistema formado pela união de dois subsistemas estatisticamente independentes (A e B), sendo  $S_q(A)$  e  $S_q(B)$  a  $q$ -entropia desses subsistemas, respectivamente. Em tal situação, são descritos pelas distribuições  $P_A$  e  $P_B$ , respectivamente, as probabilidades são decompostas na forma de produto  $P = P_A * P_B$  e, daí se verifica que essa  $q$ -entropia tem o que é chamado grau da não extensividade [8] que está regulada pelo índice  $q$ . Essa pseudo-aditividade é dada por  $S_q(A+B) = S_q(A) + S_q(B) + (1-q)S_q(A)S_q(B)$ . Nesse contexto, o chamado índice entrópico  $q$  mede o grau da não extensividade, onde para  $q < 1$  corresponde ao caso de superextensividade,  $q = 1$  extensividade e  $q > 1$  subextensividade. Tanto para superextensividade quanto para subextensividade tem-se  $S_q \geq 0$ .

Neste artigo, foi utilizada a versão contínua da entropia  $S_q$ , com função de densidade de probabilidade  $P$ , o funcional da entropia não extensiva de Tsallis [19] fica da forma

$$S_q = k \frac{1 - \int [P(x)]^q dx}{q - 1} \quad (3)$$

onde  $q$  é interpretado como parâmetro de não extensividade que depende do alcance da interação ou quão longe do equilíbrio o sistema se encontra. Utilizando o formalismo matemático dos multiplicadores de Lagrange para maximizar o funcional entrópico (3) num processo de otimização global [7,17] pode-se obter a  $q$ -generalização da gaussiana como solução otimizada. Assim, como distribuições gaussianas estão conectadas à entropia BGS, as  $q$ -gaussianas estão conectadas à entropia  $S_q$  da seguinte forma

$$P_q(x) = A \exp_q(-\beta x^2) \equiv A [1 - \beta(1-q)x^2]^{-\frac{1}{1-q}} \quad (4)$$

sendo  $A$  uma constante de normalização e  $\beta$  o parâmetro de Lagrange correspondente. Essa função para  $q > 1$  decai assintoticamente como uma lei de potência [15,16]

$$P_q(x) \approx x^{-\frac{2}{q-1}}. \quad (5)$$

A entropia de Tsallis é uma função côncava para todo o  $q > 0$  e convexa para  $q < 0$  e, portanto, entropia de um sistema em equilíbrio é um extremo [15,16].

### 3. RESULTADOS E DISCUSSÃO

Com base nos dados do PDB das famílias, utilizamos o programa SizPDB, desenvolvido pelo grupo, em linguagem C, para extrair a massa e o raio de giração de todas as 1213 proteínas. O comportamento entre o raio de giração e a massa das proteínas apresenta uma lei de potência [6]. Nesse caso, apresenta um comportamento fractal que pode sugerir a compactação dessas macromoléculas.

Aqui nos propomos medir a densidade média de compactação, quer de forma grosseira, como  $\rho_j = M_j / V_j$ , onde  $M_j$  é a massa da  $j$ -ésima proteína e  $V_j$  é o volume da esfera definido pelo raio de giro.

Esboçamos na Figura 1 o comportamento da densidade média de compactação em função do raio de giro das proteínas.

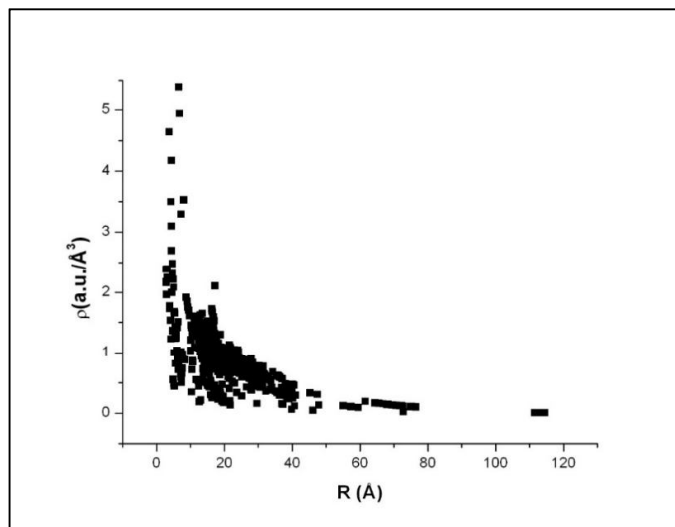


Figura 1: Comportamento da densidade média de compactação de 1213 proteínas. Fonte: Elaborado pelos autores, com base na pesquisa realizada.

A densidade tende à média de compactação no valor de  $\langle \rho \rangle = 0,86 \text{ u.m.a.} / \text{Å}^3$ , e cuja densidade média se comporta como uma distribuição normal, como mostra a Figura 2, coincidindo com a densidade média do trabalho de Moret e colaboradores [7], mas com desvio padrão 0.7.

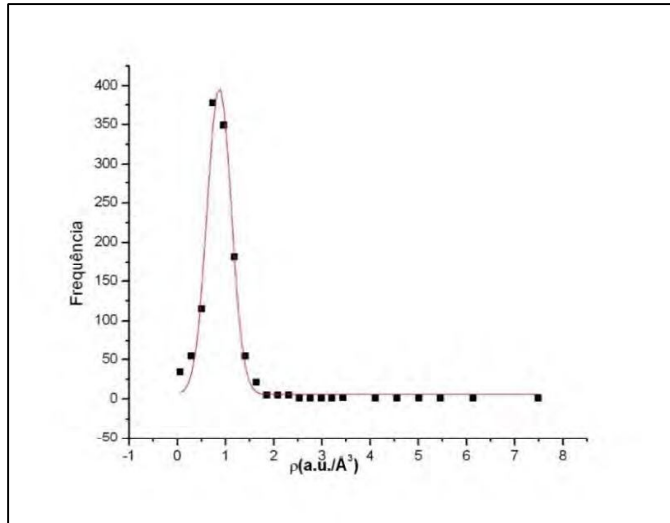


Figura 2: Distribuição normal da densidade média de compactação com  $\langle \rho \rangle = 0,86 \text{ u.m.a.} / \text{Å}^3$ ,  $\sigma = 0,7$  e  $\text{erro} = 0,01 \text{ u.m.a.} / \text{Å}^3$ . Fonte: Elaborado pelos autores, com base na pesquisa realizada.

Portanto, este estudo mostra que o enovelamento das proteínas de repetição em tandem tem uma densidade uniforme de compactação, em função do raio de giração.

A Figura 3 mostra a distribuição das massas das proteínas adapta-se tanto a distribuição gaussiana quanto a q-gaussiana, para os dados experimentais. No entanto, é interessante notar que, mesmo com esse bom ajuste as curvas são estatisticamente diferentes pelo F-valor.

10 e 11 de setembro de 2015 / Salvador, Bahia, Brasil

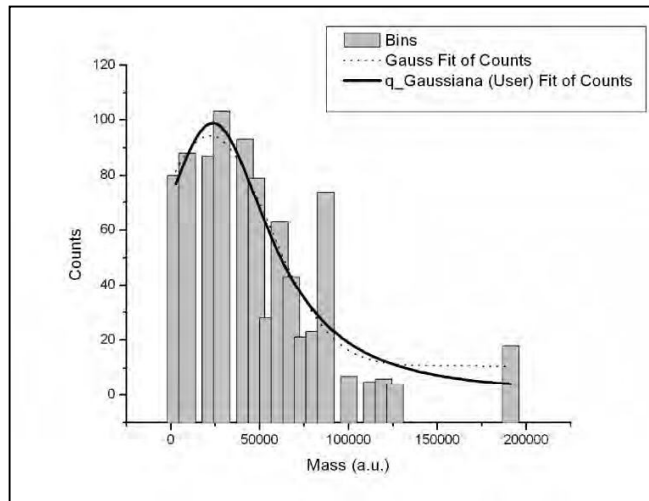


Figura 3. A distribuição  $q$ -Gaussiana (linha azul) das 1213 massas de proteínas de repetição em tandem com índice entrópico  $q = 1.90 \pm 0,83$ , a distribuição de Gauss (linha pontilhada) ambos com coeficientes de correlação de Pearson  $\approx 0,84$  e o valor  $F \approx 51$  indicando que os dois conjuntos de dados são estatisticamente diferentes. Fonte: Elaborado pelos autores, com base na pesquisa realizada.

Massas e volumes de proteína são variáveis extensivas. Neste sentido, o comportamento  $q$ -Gaussiano observado na Figura 3 é devido à auto-similaridade que é uma característica inerente deste tipo de sistema complexo, da Equação 5 a distribuição segue uma lei de potência sugerindo que o dobramento de proteínas se comporta como um objeto SOC.

#### 4. CONCLUSÃO

Ao aplicar o modelo da distribuição  $q$ -gaussiana em 1213 massas das proteínas de repetição em tandem foi possível associar, de um estado enovelado, a ideia de um processo recursivo de um sistema dinâmico. Aqui, a estrutura primária pode ser tratada como um análogo ao ponto inicial e o processo de dobra como regra recursiva, o caminho aleatório de dobragem como órbita e, finalmente, a dobra nativa como um fractal.

Em resumo, volumes e massas de proteínas de repetição em tandem se comportam como objetos fractais. As massas das proteínas seguem uma distribuição  $q$ -Gaussiana, com índice entrópico observado  $q < 3$ . Da equação 5, a distribuição das massas e volumes das proteínas de repetição seguem uma lei de potência e, portanto, o enovelamento de proteínas com repetição em tandem pode ser visto como um sistema SOC.

Em futuros trabalhos, a utilização das equações de Fokker-Planck, cuja solução estacionária é uma distribuição  $q$ -gaussiana pode levar a novas simulações de um sistema SOC para o enovelamento das proteínas fibrosas [6,5].

## 5. REFERÊNCIAS

- [1] Karplus, M. & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. **Nature Struct. Biol.** 9, 646–652.
- [2] Torrens, F. Fractal Hybrid Orbitals Analysis of the Tertiary Structure of Protein Molecules. **Molecules** 2002, 7, 26-37
- [3] F. Torrens, Análisis fractal de la estructura terciaria de las proteínas. **Encuentros en la Biología**, ISSN-e 1134-8496, Nº. 64, 2000.
- [4] T. Vicsek, Fractal Growth Phenomena, Singapore: **World Scientific**, 1989.
- [5] M. A. Moret, J. G. V. Miranda, E. Nogueira, Jr., M. C. Santana and G. F. Zebende. Self-similarity and protein chains. **Phys. Rev. E** 71 (2005).
- [6] F. S. Silva, M. A. Moret e M. C. Santana. Análise fractal de proteína com repetição em tandem. **XVIII Encontro Nacional de Modelagem Computacional**. [No prelo] (2015).
- [7] Moret, M.A.; Santana, M.C.; Nogueira, E. e Zebende, G.F. Protein chain packing and percolation threshold. **Physica A: Statistical Mechanics and its Applications**, v.361, no.1, 2006 Feb 15, p.250(5).
- [8] Bak, P. How Nature Works: The Science of Self-Organized Criticality. **Springer-Verlag**, 1996.
- [9] B. B. Mandelbrot **The Fractal Geometry of Nature** W. H. Freeman and Co., New York (1982)
- [10] A. R. Plastino and A. Plastino, Non-extensive statistical mechanics and generalized Fokker-Planck equation, **Physica A** 222, 347 (1995).
- [11] D. Voet and J. G. Voet. **Bioquímica**. 4. ed. Porto Alegre, Artmed, 2013.
- [12] Andrey V, K. Tandem repeats in proteins: from sequence to structure. **J Struct Biol.** 179(3):279-88 (2012).
- [13] P. Enkhbayar, H. and R. H. Kretsinger. Helical parameters and correlations of tandem Leucine Rich Repeats in proteins. **Journal of Proteomics & Bioinformatics**; 7(6):139-150 (2014).



[14] M. A. Andrade, C. Perez-Iratxeta and C. P. Ponting. Protein Repeats: Structures, Functions, and Evolution. **Journal of Structural Biology** 134, 117–131 (2001).

[15] C. Tsallis, J. **Stat. Phys.** 52 (1988) 479-487.

[16] K. K. Jernigan and S. R. Bordenstein. Tandem-repeat protein domains across the tree of life. *PeerJ*. 3 (2015).

[17] C. Tsallis. The Nonadditive Entropy  $S_q$  and Its Applications in Physics and Elsewhere: Some Remarks. **Entropy** (2011).

[18] L. J. Sham, W. Kohn, **Phys. Rev.** 145 (1966) 561.

[19] C. Tsallis, **Introduction to Nonextensive Statistical Mechanics** - Approaching a Complex World. Springer (2009).