Available online at www.sciencedirect.com
ScienceDirect
Physica A 387 (2008) 2019–2024
PHYSICA A
www.elsevier.com/locate/physa

Dihedral-angle Gaussian distribution driving protein folding☆

P.H. Figueirêdo[a], M.A. Moret[b,c], E. Nogueira Jr.[d], S. Coutinho[a,*]

[a] *Departamento de Física, Universidade Federal de Pernambuco, CEP 50670-901, Recife, Pernambuco, Brazil*
[b] *Programa de Modelagem Computacional, Fundação Visconde de Cairu, CEP 40226-900, Salvador, Bahia, Brazil*
[c] *Departamento de Física, Universidade Estadual de Feira de Santana, CEP 44031-460, Feira de Santana, Bahia, Brazil*
[d] *Instituto de Física, Universidade Federal da Bahia, CEP 40210-340, Salvador, Bahia, Brazil*

Received 19 November 2007
Available online 28 November 2007

Abstract

The proposal of this paper is to provide a simple angular random-walk model to build up polypeptide structures, which encompass properties of dihedral angles of folded proteins. From this model, structures will be built with lengths ranging from 125 up to 400 amino acids for the different fractions of secondary structure motifs, in which dihedral angles were randomly chosen according to narrow Gaussian probability distributions. In order to measure the fractal dimension of proteins three different cases were analyzed. The first contained α-helix structures only, the second β-strands structures and the third a mix of α-helices and β-sheets. The behavior of proteins with α-helix motifs are more compact than in other situations. The findings herein indicate that this model describes some structural properties of a protein and suggest that randomness is an essential ingredient but proteins are driven by narrow angular Gaussian probability distributions and not by random-walk processes.

PACS: 87.14.Ee; 87.15.Aa; 05.45.Df; 05.40.Jc

Keywords: Protein structure; Angular Gaussian-walk; Radius of gyration; Mass fractal exponent

The manner in which a protein folds from a random coil into a unique native state in a relatively short time is one of the fundamental puzzles of molecular biophysics. It is well accepted that a unique native three-dimensional structure, characteristic of each protein and determined by the sequence of its amino acids' sequence, dictates protein functions. The folding process should involve a very complex molecular recognition phenomenon depending on the interplay of many relatively weak non-bonded interactions. This would lead to a huge number of possible final conformations under conventional molecular optimization methods based on the search for the minima of the energy hyper-surface. This number, which should increase with the number of the chain's degrees of freedom, however, is severely restricted during the real folding process, excluding relevant portions of the energy landscapes as far as an extended or random conformation is chosen as the initial state [1–10]. On the other hand, if the extreme limit, were considered, where a polypeptide chain departs from its denatured state and in relatively very short period of time finds its unique native

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

* Corresponding author. Tel.: +55 81 21267633; fax: +55 81 32710359.
E-mail addresses: phf@lftc.ufpe.br (P.H. Figueirêdo), moret@cairu.br (M.A. Moret), enogue@ufba.br (E. Nogueira Jr.), sergio@lftc.ufpe.br (S. Coutinho).

0378-4371/$ - see front matter
doi:10.1016/j.physa.2007.11.034

Table 1
Seven possible pairs of dihedral angles and the associated conformations occurring in several amino acid sequences [13]

Φ	Ψ	Conformation
−65	−40	*A*
−89	−1	*C*
−117	142	*B*
−69	140	*P*
78	20	*G*
103	−176	*E*
−83	133	*O*

The α-helix pair is denoted by A while the β-strand pair is denoted by B.

state after searching amongst the astronomical number of possible configurations, the simulating process for proteins with fifty to five hundred amino acids using approaches such as Monte Carlo and molecular dynamics, becomes impracticable, due to the very high computation cost. Such a contradictory dynamical picture is known as the *Levinthal paradox* [11].

To investigate the role of stochasticity on the final native state, an inverse strategy is proposed, based on a simple angular three-dimensional random-walk model to build up protein backbones with different lengths and distinct percentages of secondary structures. In the proposed model, each step has a fixed radial size l_0 but dihedral Φ and Ψ angles of the protein backbone are chosen according to independent Gaussian probability distributions, following the suggestion given in Ref. [12]. The mean value and standard deviation of each is defined according to the allowed regions of the Φ/Ψ plot of the frequency distribution of dihedral angles, the so-called Ramachandran map. Φ and Ψ mean values were used as proposed in the PRELUDE software package [13]. These values were computed from comparative statistics of the backbone secondary structure for several amino acid sequences. Table 1 indicates the seven possible pairs of (Φ, Ψ) dihedral angles and the associated structures of the main chain backbone, as predicted by this method. These specific angles describe the average conformation of a wide range of proteins with known backbone structures. To simulate structures with a definite percentage of secondary structures f, a characteristic number of steps n is fixed and the growth process within these n steps is divided into two stages:
(1) The first $n \times f$ steps are built according to an angular Gaussian probability distribution, whose mean value is one pair of angles as seen in Table 1, which in turn is associated with a given structure.
(2) The next $n \times (1 - f)$ steps are built according to an angular Gaussian probability distribution, whose mean value at each step is randomly chosen from amongst the seven pairs of angles of Table 1.

For the following n steps rules 1 and 2 are repeated upwards to construct a peptide chain with N amino acids. Therefore, in order to obtain an appropriate choice of the f percentage this stochastic procedure assures that the final peptide main chain follows the Ramachandran map.

Within this simple model structures of the protein backbone were constructed considering only the dihedral angles. All other bonded or non-bonded interactions were not explicitly considered as well as excluded volume and steric effects, which are expected to be taken into account by the appropriate choice of the average values of the Gaussian probability distributions. For this reason it was possible to generate an elevated number of samples of possible protein conformations. An exhaustive number of simulations ($\sim 10^4$) were performed considering three basic cases: (a) $f = 0.6$ with α-helix structures; (b) $f = 0.6$ with β-strand structures and (c) the first $n/2$ steps built with $f = 0.6$ of α-helix structures and the next $n/2$ steps with $f = 0.6$ of β-strand structures, consecutively. Therefore, in the α-case (a) 60% of the amino acids corresponds on average to α-helix structures, in the β-case (b) 60% of the amino acids corresponds to β-strands, while in the mixed-case (c) the whole structure has an average of 30% of α-helix and 30% of β-strands. 10^4 chains of the total size varying from $N = 125$–400, with the number of steps $n = 100$ were generated. For each case described above there was a variation of f in the interval [0, 1], step 0.1. There was also a variation of the standard deviation σ of the Gaussian distribution within a wide range of values from 0 to π.

Fig. 1 shows the average radius of gyration ($\langle R_g \rangle$) as a function of the number of amino acids (N) for the three distinct choices of structures. From this plot however, a power-law behavior pattern can be observed indicating that these structures are self-similar. The corresponding scaling exponent, which somehow describes the compactness of the structure, is calculated by the scaling relation: $R_g \sim N^{\nu}$, in all cases. The characteristic scaling exponents are $\nu = 0.401 \pm 0.002$ for the α-helix case and $\nu = 0.417 \pm 0.002$ for the β-strands case, with these values falling in