

“Fuzzy oil drop” model verified positively

Mateusz Banach^{1,2}, Katarzyna Prymula^{1,3}, Leszek Konieczny^{1,4}, Irena Roterman^{1*}

¹Department of Bioinformatics and Telemedicine, Medical College, Jagiellonian University, Lazarza 16, 31-530 Krakow, Poland; ²Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Reymonta 4, 30-059 Krakow, Poland; ³Faculty of Chemistry, Jagiellonian University Ingardena 3, 30-060 Krakow, Poland; ⁴Chair of Medical Biochemistry, Medical College, Jagiellonian University, Kopernika 7, 31-034 Krakow, Poland; Irena Roterman - Email: myroterm@cyf-kr.edu; Phone: +48 12 619 96 93; *Corresponding author

Received December 19, 2010; Accepted December 29, 2010; Published February 07, 2011

Abstract:

The “fuzzy oil drop” model assuming the structure of the hydrophobic core of the form of 3-D Gauss function appeared to be verified positively. The protein 1NMF belonging to downhill proteins was found to represent the hydrophobic density distribution accordant with the assumed model. The accordance of the protein structure with the assumed model was measured using elements of theory information. This observation opens the possibility to simulate the folding process as influenced by external force field of hydrophobic character.

Keywords: hydrophobic core, downhill proteins

Background:

The protein molecule was suggested by Kauzmann to represent the form of oil drop [1]. This model which is treated as discrete one was extended to the form of “fuzzy oil drop” taking the 3-D Gauss function to describe the idealized hydrophobicity density distribution [2]. The highest hydrophobicity concentration is expected in the center of the protein body with the decrease of its values toward the surface where the hydrophobicity is expected to be close to zero. This is the way to express the hydrophobic character of protein molecule particularly in respect to its tertiary structure stabilization.

The assumption of the regular accordant with 3-D Gauss function appeared not appropriate for proteins representing the significant irregularity versus assumed model in the biological function-related area [3-7]. The protein 1NMF was selected in this paper to represent the structure accordant with the assumed “fuzzy oil drop” model.

Methodology:

Data The protein belonging to downhill proteins – 1NMF was selected for analysis [8]. This is the protein of the β -barrel structure. This protein belongs to the group of cold shock proteins.

“Fuzzy oil drop” model

The idealized hydrophobicity density distribution in protein body is assumed to be represented by 3-D Gauss function (see **supplementary material 1**). This is why these values can be considered equal to zero. The size of the molecule is expressed by the triple $\sigma_x, \sigma_y, \sigma_z$, which is calculated for each molecule individually provided that the orientation of the molecule with the longest possible inter-effective atoms distance is determined according to the appropriate coordinate system axis. The σ values are calculated as the 1/3 of the longest distance between two effective atoms calculated along each axis. The value of the Gauss function at any point of protein body is treated as the idealized hydrophobic density defining the hydrophobic core.

The idealized hydrophobicity at any point of the “fuzzy oil drop” can be calculated according to the Gauss function for the molecule located with its geometric center as the origin of the coordinate system.

On the other hand, The empirical hydrophobicity distribution is calculated according to the function presented by Levitt [9]. (See **supplementary material 2**)

Kullback-Leibler information entropy:

The accordance between the idealized and the observed hydrophobicity distribution is measured according to the Kullback-Leibler relative (divergence) entropy [11], which quantifies the distance between two distributions. The distance between the observed and the theoretical (O/T) distribution was calculated. This value can be estimated only with respect to other solutions. The random distribution of hydrophobicity represented the border case for which the distance (O/R) was calculated. The relation $O/T < O/R$ was taken as evidence for a non-random distribution close to theoretical one (see **Supplementary material 3**)

Discussion:

Structure interpreted using the “fuzzy oil drop” model:

The results of D_{KL} calculation are as follows: O/T equal to 0.1715 and O/R equal to 0.5832. The relation $O/T < O/R$ and low value of O/T suggests the hydrophobicity distribution similar to the theoretically expected. Large value of O/R supports this observation.

The profile for 1NMF showing the hydrophobic interactions collected by effective atoms of each residue as the effect of interactions with other amino acids is shown in **Figure 1A**.

The 3-D presentation of protein 1NMF with residues (marked in white) with strongest hydrophobic interactions (responsible for the generation of the hydrophobic core) is given in **Figure 1B**.

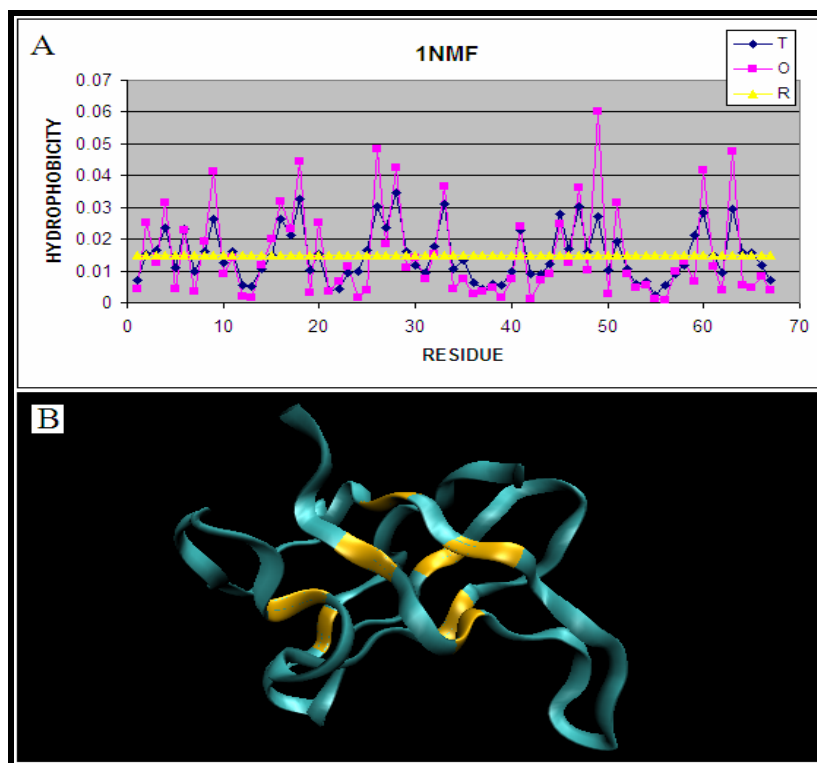


Figure 1: The hydrophobic density distribution in 1NMF. **A:** profile of idealized (T), observed (O) and random (R) distributions. **B:** 3-D presentation of the 1NMF structure with residues responsible for hydrophobic core generation distinguished in yellow (maxima on the O profile shown in A)

Conclusions:

The proteins folded following the “fuzzy oil drop” model are expected to represent the molecules very well soluble with no biological activity. Entire coverage of the protein surface with hydrophilic residues could result as no tendency to interact with any other molecule. Taking this conclusion into consideration no protein satisfying the condition of idealized hydrophobic core structure was expected. Although the proteins of biological function requires to be highly soluble with no specific interaction were found. The antifreeze proteins seem to represent the biological activity of this category [12]. The downhill proteins, experimentally proved as folding very fast seem to be influenced by water environment directing the hydrophobic residues toward the center of protein body with simultaneous exposure of hydrophilic residues on the protein surface [13]. The cold shock proteins (the protein under consideration belongs to this group of proteins) is oriented on non-specific RNA binding keeping RNA unfolded. The protein appears in the stress condition (temperature stress). This is why is expected to appear and fold

When folding, the selected protein satisfies all the conditions defined by non-bonding interactions with simultaneous hydrophobic core formation. Hydrophobic residues located in the central part of the molecule and exposure of hydrophilic residues on the surface are the main tenets of the “oil drop” model introduced by Kauzmann [1]. The Kullback-Leibler entropy [11], which is a measure of the distance between the target distribution (idealized one) and the observed one in a particular molecule revealed good accordance of the observed hydrophobicity distribution with the idealized one.

Although the protein described in this paper as accordant with the “fuzzy oil drop” model is of the category “easy predictable” (according to CASP classification [14] the meaning of the presented model expresses its possible applicability to folding process simulation taking into account the influence of external force field.

References:

- [1] W Kauzmann, *Adv Protein Chem* **14**: 1 (1959)
- [2] L Konieczny *et al. In Silico Biol* **6**: 15 (2006) [PMID: 16789910]
- [3] M Bryliński *et al. Comp. Biol. Chem* **30**: 255 (2006a) [PMID: 16798094]
- [4] M Brylinski *et al. Biochimie* **88**: 1229 (2006b) [PMID: 16647798]
- [5] M Bryliński *et al. J. Biomol Struct. Dynam* **23**: 519 (2006c) [PMID: 16494501]
- [6] M Bryliński *et al. PLoS-Computational Biology* **3**: e94 (2007) [PMID: 17530916]
- [7] M Brylinski *et al. Int J Bioinform Res Appl* **3**:234 (2007) [PMID: 18048191]
- [8] A Schnuchel *et al. Nature* **364**: 169 (1993) [PMID: 8321289]
- [9] M Levitt, *J Mol Biol* **104**: 59 (1976) [PMID: 957439]
- [10] A Aboderin *Int J Biochem* **2**: 537 (1971)
- [11] R Nalewajski Amsterdam *et al.* Elsevier, 2006.
- [12] Banach *et al.* submitted
- [13] Banach *et al.* submitted
- [14] C Orengo *et al. Proteins. Suppl* **3**: 149 (1999) [PMID: 10526364]

Edited by P. Kanguane

Citation: Banach *et al.* Bioinformatics 5(9): 375-377 (2011)
distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

Supplementary material:

Material 1:

The idealized hydrophobicity density distribution in protein body is assumed to be represented by 3-D Gauss function:

$$\tilde{H}t_j = \frac{1}{\tilde{H}t_{sum}} \exp\left(\frac{-(x_j - \bar{x})^2}{2\sigma_x^2}\right) \exp\left(\frac{-(y_j - \bar{y})^2}{2\sigma_y^2}\right) \exp\left(\frac{-(z_j - \bar{z})^2}{2\sigma_z^2}\right)$$

where $\bar{x}, \bar{y}, \bar{z}$ are the coordinates of the geometric center of the molecule (usually located in the origin of the coordinate system).

Material 2:

The empirical hydrophobicity distribution is calculated according to the function presented by Levitt [9].

$$\tilde{H}o_j = \frac{1}{\tilde{H}o_{sum}} \sum_{i=1}^N (H_i^r + H_j^r) \begin{cases} \left[1 - \frac{1}{2} \left(7 \left(\frac{r_{ij}}{c} \right)^2 - 9 \left(\frac{r_{ij}}{c} \right)^4 + 5 \left(\frac{r_{ij}}{c} \right)^6 - \left(\frac{r_{ij}}{c} \right)^8 \right) \right] & \text{for } r_{ij} \leq c \\ 0 & \text{for } r_{ij} > c \end{cases}$$

where N expresses the number of amino acids in the protein (number of grid points), \tilde{H}_i^r expresses the hydrophobicity of the i -th residue according to the accepted hydrophobicity scale (the Aboderin scale was applied in this work [10], r_{ij} expresses the distance between the i -th and j -th interacting residues, and c expresses the cutoff distance, which according to the original paper [10] is assumed to be 9 Å. The values of $\tilde{H}o_j$ are standardized by

dividing them by the coefficient $\tilde{H}o_{sum}$, which is the sum of all hydrophobicities attributed to grid points.

Material 3:

The relation $O/T < O/R$ was taken as evidence for a non-random distribution close to theoretical one.

$$D_{KL}(p|p^0) = \sum_{i=1}^N p_i \log_2(p_i / p_i^0),$$

where: D_{KL} – distance entropy, p – probability of a particular observed event, p^0 – probability in reference distribution. The theoretical (T) and random (R) distributions were taken for independent calculations as reference distributions for observed (O) one in protein. The index “ i ” denotes a particular amino acid. N denotes the number of amino acids in the polypeptide chain.