

Fuzzy-Oil-Drop Hydrophobic Force Field – A Model to Represent Late-stage Folding (*In Silico*) of Lysozyme

<http://www.jbsdonline.com>

Michal Brylinski^{1,2}
Leszek Konieczny³
Irena Roterman^{1,4,*}

¹Department of Bioinformatics
and Telemedicine
Collegium Medicum –
Jagiellonian University
Kopernika 17, 31-501 Krakow, Poland
²Faculty of Chemistry
Jagiellonian University
Ingardena 3, 30-060 Krakow, Poland
³Institute of Medical Biochemistry
Collegium Medicum –
Jagiellonian University
Kopernika 7, 31 034 Krakow, Poland
⁴Faculty of Physics
Jagiellonian University
Reymonta 4, 30-060 Krakow, Poland

Abstract

A model of hydrophobic collapse (*in silico*), which is generally considered to be the driving force for protein folding, is presented in this work. The model introduces the external field in the form of a *fuzzy-oil-drop* assumed to represent the environment. The drop is expressed in the form of a three-dimensional Gauss function. The usual probability value is assumed to represent the hydrophobicity distribution in the three-dimensional space of the virtual environment. The differences between this idealized hydrophobicity distribution and the one represented by the folded polypeptide chain is the parameter to be minimized in the structure optimization procedure. The size of *fuzzy-oil-drop* is critical for the folding process. A strong correlation between protein length and the dimension of the native and early-stage molecular form was found on the basis of single-domain proteins analysis. A previously presented early-stage folding (*in silico*) model was used to create the starting structure for the procedure of late-stage folding of lysozyme. The results of simulation were found to be promising, although additional improvements for the formation of β -structure and disulfide bonds as well as the participation of natural ligand in folding process seem to be necessary.

Introduction

The idea of hydrophobic interaction and the hydrophobic core or nucleus was introduced almost as soon as the protein folding problem appeared in biological research (1). It is generally accepted that globular proteins consist of a hydrophobic core and a hydrophilic exterior (2-7). Estimation of the distribution of packing density in general (8) and of the hydrophobicity distribution that lead to the creation of the hydrophobic core in the protein molecule appeared to be the criterion for predicted protein structure in *ab initio* approaches (9). A comparison of the hydrophobic core in proteins to an 'oil-drop' (1) is the basis of the model presented in this paper. The discrete form of 'hard' core and 'soft' interacting surfaces introduced by Klapper (3) is changed to a continuous *fuzzy-oil-drop*, with the hydrophobic center of highest hydrophobicity value localized in the center of the drop, and with a distance dependent decrease of hydrophobicity according to Gauss function. The hydrophobicity distribution in a virtual force field reaches its maximum (the gaussian function value for a common mean value) in the center of the ellipsoid (0,0,0) and then decreases to negligibly small values on the surface of the ellipsoid.

The external force field introduced in this paper represents the environment for protein folding. The distribution of hydrophobicity according to a gaussian function (the probability values are treated here as hydrophobicity values) represents the idealized virtual hydrophobic space. The folding polypeptide is aimed at creating the hydrophobic core, which is more or less similar to the expected one. The grid points represent the hydrophobicity values according to a gaussian function, and simultaneously reveal the values of hydrophobicity resulting from the interacting

*Email: myroterm@cyf-kr.edu.pl

residues. The optimization procedure is aimed at minimization of the differences between these two representations.

Highly desirable databases including information as to correlation between protein chain length and molecular dimension, distribution of the distances between residues exposed to the solvent, as well as other aspects might be useful for structure prediction. The relation between polypeptide chain length and the size of the final molecular form as well as the molecular density was found for a number of conformational states of protein molecules (10, 11). The size of *fuzzy-oil-drop* is critical for the folding process in this model. The size change between the early-stage folding structural form and the native one was estimated quantitatively on the basis of single-domain proteins. The dependence of drop size and radius of gyration on the number of amino acids for both conformational states (native and early-stage) was revealed to deliver the common pattern for any polypeptide *fuzzy-oil-drop* size change during the folding process *in silico*.

The model of the early-stage folding structural form of the polypeptide has been described elsewhere in terms of geometry (12) and information-theory (13). The limited conformational sub-space was applied to BPTI (14), lysozyme (15), ribonuclease (13) and the α and β chains of hemoglobin (16). The evaluation of the early-stage structural forms in respect of the presence of biological function-related structural motifs was presented in (17). The generalization of sequence-to-structure dependence (structure understood as early-stage form) expressed in the form of a contingency table created a tool to construct the early structure of any polypeptide (18). The coding system for structural motif classification of early-stage folding structures, together with the information stored in the sequence-to-structure contingency table, enables quantitative estimation of the degree of difficulty of structure prediction for any amino acid sequence (19).

All these models taken together constitute the preliminary step of the late-stage model presented in this paper.

Materials and Methods

Data

The lysozyme molecule, PDB ID: 2EQL (20), was taken as the model. The following structural forms of lysozyme were analyzed in this paper: native, early-stage (initial), and late-stage (final). The early-stage structure of lysozyme taken as initial for the late-stage folding model presented in this paper was predicted from the amino acid sequence according to a previously described sequence-to-structure contingency table (18). The algorithm of early-stage structure prediction is described in detail in (19).

The Change of Fuzzy-Oil-Drop Size Between Early-Stage and Native Conformational State

The early-stage structural form of the polypeptide differs significantly with respect to the shape and size of the molecule. The size understood as the smallest rectangular box covering the complete molecule was calculated as follows. The side chains were simplified and represented by one effective "atom" placed in the geometrical center of the side chain. The geometrical center is defined as the middle of the distance between the two most remote atoms belonging to a particular side chain. The longest distance between two effective atoms was taken as the D_z measure (distance along the Z-axis); the longest distance between two effective atoms in the XY-plane was taken as the D_x measure (distance along the X-axis); and finally the difference between the highest and lowest values of Y was taken as the measure of the D_y box edge. Moreover, the distance between each box edge and the nearest point repre-

senting an effective atom was extended by the hydrophobic cutoff used (for reasons given in the ‘*External Hydrophobic Force Field*’ section), which has the fixed value 9.0 Å. The box volume (V) was expressed as $V = D_z \times D_x \times D_y$.

The approximation function expressing the dependence of D_z , D_x , D_y , and V on the number of amino acids in a single-domain protein is used to estimate the presumable size of the box completely covering the protein in its native form. Single-domain proteins were selected with the aid of CATH Domain Structure Database (21-23). Their three-dimensional structures (only that determined by X ray crystallography) were obtained from Protein Data Bank (24). For each protein an early-stage conformation state was created using a “step-back” unfolding procedure presented elsewhere (13-16). Three sets of parameters were calculated for both native and early-stage conformational states of the proteins:

I. The box volume covering the complete molecule: $V = D_z \times D_x \times D_y$,

II. The ratio of the box edges expressed as $D_z : D_x : D_y$,

III. The radius of gyration (R_g), using the following equation (25):

$$R_g = \sqrt{\frac{1}{(N+1)^2} \sum_{i=1}^{N-1} \sum_{j=1}^N \langle (\vec{r}_i - \vec{r}_j)^2 \rangle} \quad [1]$$

where N is the chain length, \vec{r}_i and \vec{r}_j are the coordinates of C α atom of i -th and j -th residue, respectively.

External Hydrophobic Force Field and Optimization Procedure

It is generally accepted that the second step of folding is driven mainly by hydrophobic interactions (4, 26-30). The new hydrophobic force field in the form of a *fuzzy-oil-drop* is assumed to represent the environment for the folding protein molecule. The early-stage structure of a polypeptide taken as initial for the late-stage folding simulation is placed in a box as described in previous section. The box is filled with an internal three-dimensional grid similar to the ones used in lattice models (31-36). A constant 5 Å grid size was chosen to split the difference between computational time and accuracy. Each j -th grid point is characterized by the \tilde{H}_j value, which represents the degree of hydrophobicity of the *fuzzy-oil-drop* according to three-dimensional gaussian function:

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

where σ_x , σ_y , σ_z denote standard deviations and point $(\bar{x}, \bar{y}, \bar{z})$ represents the highest hydrophobicity value and keeps a fixed position at the center of the box (0,0,0) during the simulation. The value attributed to each grid point is calculated according to the standardized gaussian function (the sum of all hydrophobicity values on grid points is equal to 1.0). The three variables (x_j, y_j, z_j) present in the three-dimensional gaussian function represent the Cartesian coordinates of particular grid point.

Each grid point is also described by $\tilde{H}o_j$, where o_j denotes the observed hydrophobicity and j identifies a particular grid point. The $\tilde{H}o_j$ value expresses quantitatively the influence of residues on particular j -th grid point. The side chains were simplified and represented by one effective atom placed at the geometrical center, defined as the middle of the distance between two the most remote atoms. The function expressing the hydrophobic interaction was taken according to Levitt (37):

$$\tilde{H}o_j = \frac{1}{\tilde{H}o_{sum}} \sum_{i=1}^N H_i^r \left\{ \begin{array}{l} 1 - \frac{1}{2} \left(7 \left(\frac{r_{ij}}{c} \right)^2 - 9 \left(\frac{r_{ij}}{c} \right)^4 + 5 \left(\frac{r_{ij}}{c} \right)^6 - \left(\frac{r_{ij}}{c} \right)^8 \right) \\ \text{otherwise } 0 \end{array} \right\} \text{ for } r_{ij} \leq c, \quad [3]$$

where N is the total number of residues in the protein under consideration, \tilde{H}_i denotes the hydrophobicity of the i -th residue according to the scale (transformed to the form covering the range from 0 (minimum) to 1 (maximum)) of hydrophobicity for amino acids based on the *fuzzy-oil-drop* model (submitted for publication), r_{ij} denotes the separation of the j -th grid point and the effective atom of the i -th residue, and c denotes the hydrophobic cutoff and has the fixed value of 9 Å following the original paper (37). This means that the observed hydrophobicity of j -th point is locally formed by residues within 9 Å radius. The observed hydrophobicity distribution is also standardized to the value 1.0, therefore can be compared to the theoretical hydrophobicity described previously.

The sum of the differences between idealized \tilde{H}_t and observed \tilde{H}_o hydrophobicity values over all the grid points is the parameter minimized during the optimization procedure:

$$\Delta\tilde{H}_{tot} = \sum_{j=1}^P (\tilde{H}_t_j - \tilde{H}_o_j)^2 \quad [4]$$

where P denotes the total number of grid points at a particular step of folding simulation. $\Delta\tilde{H}_{tot}$ represents the driving force to concentrate the residues of high hydrophobicity in the center of the box and to push low-hydrophobic residues toward the peripheral part of the box.

The optimization procedure is the set of two iterated steps. Each step of traditional energy optimization in the ECEPP-based procedure using ECEPP/3 all-atom force field (38-41) is followed by a step of hydrophobicity optimization. Each step of hydrophobic interaction optimization, on the other hand, is performed at step-wise decreasing sizes of the expected *fuzzy-oil-drop*. The algorithm given by Rosenbrock (42) was applied to optimize both hydrophobicity distribution and the energy of a polypeptide during the simulation. For each amino acid ϕ , ψ dihedral angles were free to rotate, while ω angles of peptide bonds were kept fixed. During energy minimization all backbone dihedral angles as well as all χ side chain angles were able to rotate.

Structure Analysis

The early-stage (initial) and late-stage (final) structures of lysozyme were compared with the native structure according to following criteria:

- I. The number of residue-residue contacts was calculated for all structural forms and was presented as contact maps. A residue-residue contact was classified as being present when the distance between two effective atoms was lower than assumed cutoff (12 or 8 Å).
- II. The number of hydrogen bonds of different types (O(i)→H-N(j), parallel and antiparallel bridges) per 100 residues was calculated with DSSP (43).
- III. RMSD-C α per residue was calculated for early-stage and late-stage forms of lysozyme using the native form as a reference structure.
- IV. The distances between the geometric center of the molecule and sequential C α atoms ($D_{\text{center-C}\alpha}$) in the polypeptide chain were calculated. The profile of these distances for each amino acid reveals a rough degree of similarity.

Differences Between Native Structure and the Structure Folded In Silico

The method to assess the correctness of the structure received *in silico* versus the expected one is applied to reveal the differences between these two structural forms

of the protein. The value $\Delta\tilde{H}$ expressing the difference is introduced to measure and localize the area of high differences. For i -th residue $\Delta\tilde{H}_i$ is calculated as follows:

$$\Delta\tilde{H}_i = \Delta\tilde{H}t_i - \Delta\tilde{H}o_i \quad [5]$$

where $\Delta\tilde{H}t_i$ and $\Delta\tilde{H}o_i$ are the theoretical and observed values of hydrophobicity for the geometric center of i -th residue, respectively.

Results

The Molecular Dimensions of Proteins in Early-Stage and Native Conformational State

The size of rectangular box covering the protein molecule entirely in its native and early-stage conformational state as dependent on number of amino acids in polypeptide chain is shown as a logarithmic plot in Figure 1. The analysis of this presentation visualizes the degree of expected drop compression during the implosion process directing the hydrophobic residues toward the center of the molecule. Following correlations between the box volume covering the complete molecule (V) and chain length (N) were found as follows:

$$\log V = 3.5671 + 0.7725 \times \log N \text{ for native conformational state} \quad [6]$$

$$\log V = 3.0013 + 1.2271 \times \log N \text{ for early-stage conformational state} \quad [7]$$

The correlation coefficients (0.95 and 0.88 for native and early-stage conformational state, respectively) point to a strong correlation between the molecular dimension and the chain length. Moreover, the ratio of the box edges expressed as $D_z : D_x : D_y$ calculated for native and early-stage structures was found to be $1.00 : 0.85 \pm 0.08 : 0.79 \pm 0.09$ and $1.00 : 0.67 \pm 0.14 : 0.53 \pm 0.12$, respectively.

Radius of Gyration Dependence on the Chain Length

The radius of gyration (R_g) estimates the characteristic volume of the globular protein and provides quantitative information on its compactness. The variation of R_g with protein chain length (N) for a number of conformational states of the protein molecules is described by:

$$R_g = K_r N^\varepsilon, \quad [8]$$

where K_r and ε are expected to be constant for a given conformational state in a wide range of chain lengths (10). The plot in Figure 2 displays radius of gyration (R_g) as a function of chain length (N) for native and early-stage conformational states of the single-domain proteins. The best fit (correlation coefficients 0.94 and 0.76, respectively) of calculated data to Equation [8] provides the following correlations:

$$R_g = 2.76N^{0.33} \text{ for native conformational state} \quad [9]$$

$$R_g = 2.15N^{0.52} \text{ for early-stage conformational state} \quad [10]$$

Lysozyme as a Test Protein

Lysozyme is a single-domain protein that contains 129 residues with four native disulfide bonds. It must be stressed that disulfide bonds formation were excluded from late-stage folding simulation. The amino acid sequence of lysozyme was used as input for early-stage structure prediction on the basis of a sequence-to-structure contingency table (18). The Structure Predictability Index (SPI) (19) calculated for the sequence was found to be 79.2, which ranks lysozyme as a moderate target for

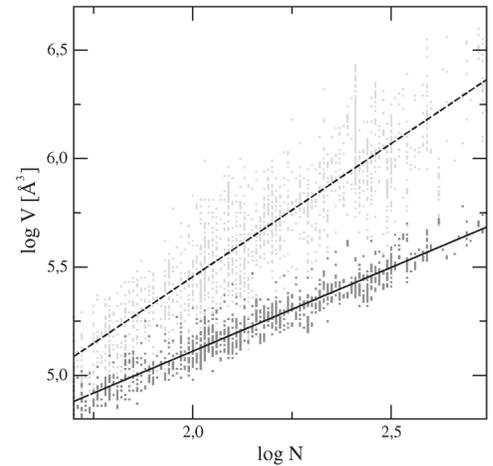
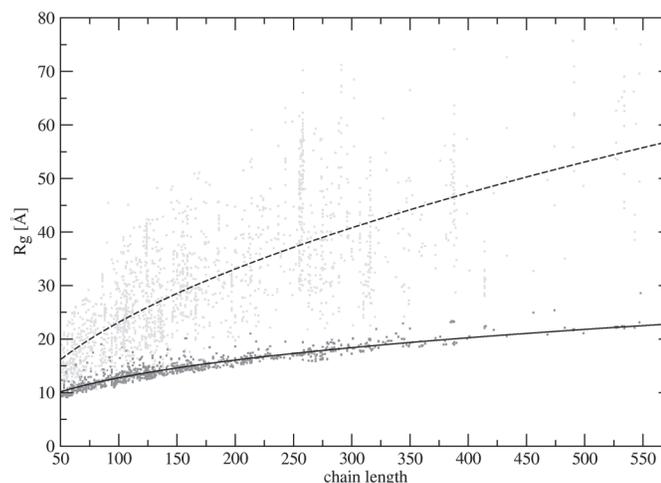


Figure 1: The size of the box completely covering protein molecule (V) in native (dark gray points) and early-stage (light gray points) conformational state versus the protein chain length (N). The profiles visualize the degree of the box (drop) compression between early-stage and final one in hydrophobic collapse simulation.

Figure 2: Variation of the radius of gyration (R_g) with the chain length for single-domain proteins in their native (dark gray points) and early-stage (light gray points) conformational states. The lines (solid and dashed for native and early-stage state, respectively) represent the best fit of the data to Equation [8].



early-stage form prediction. The Q3 (44), Q7 (19), and SOV (45, 46) parameters calculated versus native structure were found to be 62.2, 56.7, and 56.0, respectively. The predicted early-stage structural form of lysozyme was subjected to late-stage folding simulation according to *fuzzy-oil-drop* model.

Size Change of Fuzzy-Oil-Drop During Simulation

The size of the *fuzzy-oil-drop* for lysozyme expressed as $D_z \times D_x \times D_y$ was found to be $80.3 \times 70.5 \times 45.6 \text{ \AA}$ and $63.6 \times 49.5 \times 44.1 \text{ \AA}$ for early-stage and native form, respectively. Based on Equation [6] and the $D_z : D_x : D_y$ ratio calculated for native conformational state, the presumable target size of the *fuzzy-oil-drop* was predicted to be $61.7 \times 52.4 \times 48.7 \text{ \AA}$. Fairly high accordance of the target (predicted) and native (observed) sizes of the *fuzzy-oil-drop* ensured a good size-dependent condition for late-stage folding simulation. The *fuzzy-oil-drop* was linearly squeezed from the early-stage size to the predicted target size in 10 equal steps.

RMSD-C α and Residue-Residue Contacts

RMSD calculated for early-stage and late-stage structures of lysozyme using the native form as a reference structure was found to be 20.21 \AA and 17.58 \AA , respectively. Interactions that stabilize the fold are between residues that are well separated along the sequence and therefore away from the diagonal of the plot. The residue-residue contacts present in all discussed structural forms are shown in Figure 3, together with the tube representations of three-dimensional models. New residue-residue contacts appeared during late-stage folding simulation. Some of them (centered in regions: 5 vs. 125, 10 vs. 90, and 15 vs. 50) were identified as native, however the non-native residue-residue contacts (10 vs. 110, 50 vs. 125, and 90 vs. 120) were also observed.

Spatial Distribution of C α Atoms Versus the Geometrical Center

The profile of the length of vectors linking the geometrical center with sequential C α atoms ($D_{\text{center-C}\alpha}$) elucidates the three-dimensional relative displacement versus the native form of the protein (47). Profile plots are presented in Figure 4. Structural similarity may be checked by overlapping the lines representing the two compared structures. Parallel orientation of profiles is interpreted as similarity of structural forms in compared molecules oriented differently in the space. The increase of the vector length in respect to native structure is obviously due to extension of the structure, characteristic of a protein in early-stage conformational state. Similar spatial orientation of lysozyme polypeptide chain in native and late-stage conformational states can be observed for residues 1 to 45, 55 to 100, and 115 to 129. However, the profiles revealed several key resi-

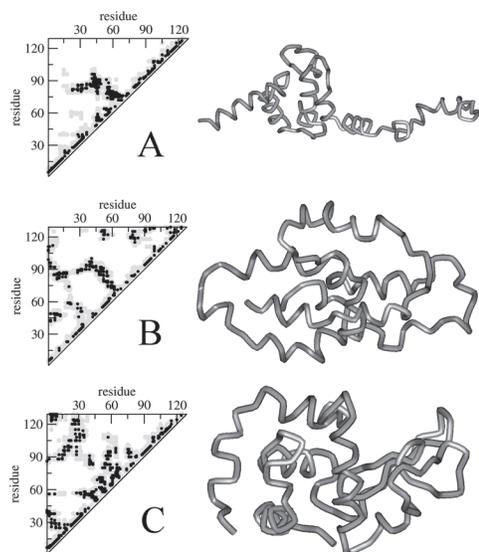


Figure 3: Residue-residue contacts in lysozyme together with tube representation of 3D models. (A) early-stage, (B) late-stage, and (C) native form. Grey and black points represent residue-residue contacts calculated using 12 and 8 \AA cutoff, respectively.

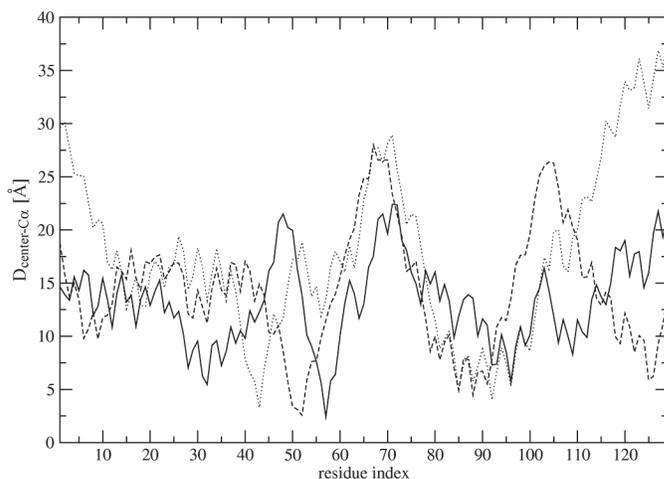


Figure 4: Spatial distribution of C α atoms versus the geometric center ($D_{\text{center-C}\alpha}$) for lysozyme in its native (solid line), early-stage (dotted line) and late-stage (dashed line) structural form.

dues, particularly residues 44, 55, 92, and 116, responsible for different arrangement of polypeptide chains in both states.

Secondary Structure Assessment

The presence of secondary structure in all discussed structural forms of lysozyme expressed as different types of hydrogen bonding is given in Table I. The *fuzzy-oil-drop* model reproduced well the helical form of polypeptide. However, the formation of β -structure remains unsatisfactory and clearly requires additional improvements.

Ligand Presence in Folding Process

The comparison of the native structure of lysozyme with the one received according to folding procedure presented in this paper reveals the aim-orientation of folding process. Figure 5 depicts the irregularities versus $\Delta\tilde{H}$ the idealized *fuzzy-oil-drop*. The main and very easily recognized difference is that in the native structure of protein the highest irregularity versus the idealized *fuzzy-oil-drop*, is localized in an active center. This means that the native structure represents the product of aim-oriented folding process.

Discussion

The protein folding is the process in which the starting structure reaches its final structural form (native) through few intermediates (31, 48-50). The commonly accepted opinion is that the starting (early-stage) step in folding is backbone conformation dependent (51-53). One of the intermediate steps is assumed to be hydrophobic collapse-dependent (54-56).

The polypeptide structure created according to the ellipse path-limited conformational sub-space was assumed to represent the early-stage of polypeptide chain (12, 13). The early-stage folding structural form of the protein was localized in a *fuzzy-oil-drop* corresponding to the protein size in its early-stage conformational state. Late-stage folding simulation consists of sequential runs of the energy optimization procedure followed by the hydrophobic interaction optimization procedure for step-wise decreasing size of the drop. The procedure stops when the size of presumable native state is reached.

The molecular density of a native form shows no changes with the chain length (10, 57) in contrast to a correlation between the apparent density and chain length observed for fully unfolded states in urea (10, 58, 59). A strong dependence of the dimension of final molecular form on the protein length was claimed to be clear (11). The relation between the molecular size (expressed as radius of gyration

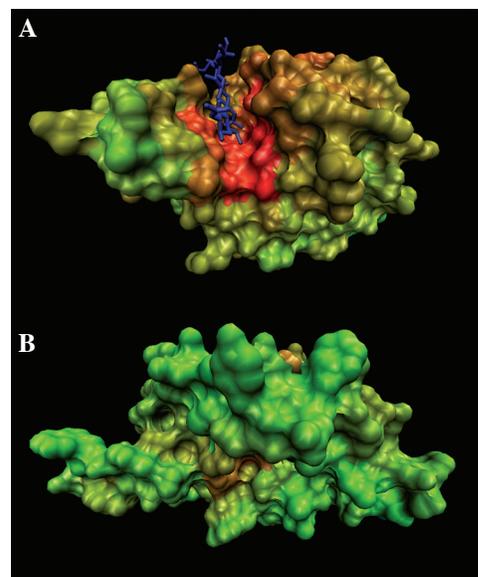


Figure 5: The distribution of hydrophobicity irregularities ($\Delta\tilde{H}$) versus the idealized *fuzzy-oil-drop*. **A**, native structure of lysozyme complexed with NAG (blue); **B**, late-stage structural form obtained as the result of late-stage folding simulation. The color scale for $\Delta\tilde{H}$ is applied (green-low to red-high difference) for three-dimensional presentation of molecule. Red color seen on a surface of native protein illustrates large irregularities ($\Delta\tilde{H}$) in the vicinity of the active site.

and box volume covering a complete protein molecule in early-stage and native conformational state) and the chain length was estimated on the basis of single-domain protein molecules. The results are in remarkably good agreement with those reported for 180 proteins in different conformational states (10). In both analyses the value of exponent ϵ in Equation [8] was found to be 0.33. Moreover, $\epsilon = 0.52$ calculated for early-stage conformational state seems to be consistent with those obtained for fully unfolded states in 8 M urea ($\epsilon = 0.52$) and 6 M GdmCl ($\epsilon = 0.54$), assuming that disulfide bonds are reduced. The dependence of the size of the box volume covering a complete protein molecule in native conformational state on the chain length together with the ratio of the box edges allows the presumable target size of *fuzzy-oil-drop* to be estimated with fairly high accuracy.

The *fuzzy-oil-drop* model for hydrophobic collapse simulation is continuous in nature. Size and shape can change elastically. The grid point outside the molecule mimics the environment of zero hydrophobicity assumed to represent the hydrophilic environment. The closer to the center of the molecule, the higher the hydrophobic density. Moreover, the early-stage folding structural form of the protein seems to work well with the model for the late-stage folding step.

The application of the late-stage model to lysozyme folding presented in this paper closely approximated the predicted structure to the native one. However, some additional improvements for example the formation of β -structure seem to be necessary. It must be noted, that disulfide bonds were absent in the simulation. Their role can be critical for folding process causing much better approach of the final structure to the native one. Moreover, the results shown in Figure 4 precisely reveal that the aim-orientation in folding process is necessary. One can assume, that the presence of a molecule mimicking the enzyme's substrate shall be present to ensure the creation of active center in a protein molecule. This assumption has been verified during the simulation of hemoglobin folding *in silico* (according to presented model) in a presence and absence of hem. The results of folding simulation of α and β polypeptide chains of hemoglobin significantly suggest that the participation of natural ligand in folding process seems to be important or even necessary.

The main advantage of the presented model is its universality, although some additional peculiarities of the folding process are expected. So far the model presented in this paper was applied to single-domain globular proteins of up to 150 amino acids long. The late-stage folding simulations of ribonuclease, BPTI, hemoglobin, and hypothetical membrane protein - target protein in CASP6 according to *fuzzy-oil-drop* model are presented elsewhere (manuscripts submitted for publication). Large multi-domain proteins including several well-separated hydrophobic cores may also be simulated using the bunch of cooperative *fuzzy-oil-drops*. Moreover the inverse function expressed as $1 - \Delta\tilde{H}_{tot}$ may be applied for "inside out" integral membrane proteins pushing the hydrophobic residues to be exposed toward the membrane. The possibility of application of the *fuzzy-oil-drop* model in such cases will be verified in close future.

Acknowledgements

Many thanks to Prof. Marek Pawlikowski (Faculty of Chemistry, Jagiellonian University) for the fruitful discussions. This research was supported by the Polish State Committee for Scientific Research (KBN) grant 3 T11F 003 28 and Collegium Medicum grants 501/P/133/L and WŁ/222/P/L.

References and Footnotes

1. W. Kauzmann. *Adv. Protein Chem.* 14, 1-63 (1959).
2. I. M. Klotz. *Arch. Biochem. Biophys.* 138, 704-706 (1970).
3. M. H. Klapper. *Biochim. Biophys. Acta.* 229, 557-566 (1971).
4. H. Meirovitch and H. A. Scheraga. *Macromolecules* 13, 1398-1405 (1980).
5. H. Meirovitch and H. A. Scheraga. *Macromolecules* 13, 1406-1414 (1980).
6. H. Meirovitch and H. A. Scheraga. *Macromolecules* 14, 340-345 (1981).

7. J. Kyte and R. F. Doolittle. *J. Mol. Biol.* 157, 105-132 (1982).
8. N. Kurochkina and G. Privalov. *Protein Sci.* 7, 897-905 (1998).
9. R. Bonneau, C. E. Strauss, and D. Baker. *Proteins* 43, 1-11 (2001).
10. O. Tcherkasskaya, E. A. Davidson, and V. N. Uversky. *J. Proteome Res.* 2, 37-42 (2003)
11. O. Tcherkasskaya and V. N. Uversky. *Proteins* 44, 244-254 (2001).
12. I. Roterman. *J. Theor. Biol.* 177, 283-288 (1995).
13. W. Jurkowski, M. Brylinski, L. Konieczny, Z. Wiiniowski, and I. Roterman. *Proteins* 55, 115-127 (2004).
14. M. Brylinski, W. Jurkowski, L. Konieczny, and I. Roterman. *Bioinformatics* 20, 199-205 (2004).
15. W. Jurkowski, M. Brylinski, L. Konieczny, and I. Roterman. *J. Biomol. Struct. Dyn.* 22, 149-158 (2004).
16. M. Brylinski, W. Jurkowski, L. Konieczny, and I. Roterman. *TASK Quarterly* 8, 413-422 (2004).
17. I. Roterman, M. Brylinski, L. Konieczny, and W. Jurkowski. *Recent Adv. In Prot. Eng.*, Ed., Sangadai, S. G. Research Signpost, Trivandrum, in press.
18. M. Brylinski, L. Konieczny, P. Czerwonko, W. Jurkowski, and I. Roterman, *J. Biomed. Biotechnol.* 2, 65-79 (2005).
19. M. Brylinski, L. Konieczny, and I. Roterman. *In Silico Biol.* 5, 0022 (2004) .
20. H. Tsuge, H. Ago, M. Noma, K. Nitta, S. Sugai, and M. Miyano. *J Biochem (Tokyo)* 111, 141-143 (1992).
21. F. Pearl, A. Todd, I. Sillitoe, M. Dibley, O. Redfern, T. Lewis, C. Bennett, R. Marsden, A. Grant, D. Lee, A. Akpor, M. Maibaum, A. Harrison, T. Dallman, G. Reeves, I. Diboun, S. Addou, S. Lise, C. Johnston, A. Sillero, J. Thornton, and C. Orengo. *Nucleic Acids Res* 33, D247-251 (2005).
22. F. M. Pearl, D. Lee, J. E. Bray, I. Sillitoe, A. E. Todd, A. P. Harrison, J. M. Thornton, and C. A. Orengo. *Nucleic Acids Res* 28, 277-282 (2000).
23. C. A. Orengo, A. D. Michie, S. Jones, D. T. Jones, M. B. Swindells, and J. M. Thornton. *Structure* 5, 1093-1108 (1997).
24. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. *Nucleic Acids Res* 28, 235-242 (2000)
25. P. J. Flory. *Principles of Polymer Chemistry*, Cornell University Press, Ithaca (1953).
26. K. A. Dill. *Biochemistry* 29, 7133-7155 (1990).
27. R. R. Matheson and H. A. Scheraga. *Macromolecules* 11, 819-829 (1978).
28. G. D. Rose, A. R. Geselowitz, G. J. Lesser, R. H. Lee, and M. H. Zehfus. *Science* 229, 834-838 (1985).
29. G. D. Rose and S. Roy. *Proc. Natl. Acad. Sci. USA* 77, 4643-4647 (1980).
30. C. N. Pace, B. A. Shirley, M. McNutt, and K. Gajiwala. *Faseb J.* 10, 75-83 (1996).
31. V. I. Abkevich, A. M. Gutin, and E. I. Shakhnovich. *Biochemistry* 33, 10026-10036 (1994).
32. J. N. Onuchic, N. D. Socci, Z. Luthey-Schulten, and P. G. Wolynes. *Fold Des.* 1, 441-450 (1996).
33. V. S. Pande and D. S. Rokhsar. *Proc. Natl. Acad. Sci. USA* 96, 1273-1278 (1999).
34. A. Sali, E. Shakhnovich, and M. Karplus. *J. Mol. Biol.* 235, 1614-1636 (1994).
35. J. Skolnick and A. Kolinski. *J. Mol. Biol.* 221, 499-531 (1991).
36. H. Taketomi, Y. Ueda, and N. Go. *Int. J. Pept. Protein Res.* 7, 445-459 (1975).
37. M. Levitt. *J. Mol. Biol.* 104, 59-107 (1976).
38. F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga. *J. Phys. Chem.* 79, 2361-2381 (1975).
39. G. Nemethy, K. D. Gibson, K. A. Palmer, C. N. Yoon, G. Paterlini, A. Zagari, S. Rumsey, and H. A. Scheraga. *J. Phys. Chem.* 96, 6472-6484 (1992).
40. L. G. Dunfield, A. W. Burgess, and H. A. Scheraga. *J. Phys. Chem.* 82, 2609-2616 (1978).
41. M. J. Sippl, G. Nemethy, and H. A. Scheraga. *J. Phys. Chem.* 88, 6231-6233 (1984).
42. H. H. Rosenbrock. *Comput. J.* 3, 175-184 (1960).
43. W. Kabsch and C. Sander. *Biopolymers* 22, 2577-2637 (1983).
44. B. Rost and C. Sander. *J. Mol. Biol.* 232, 584-599 (1993).
45. B. Rost, C. Sander, and R. Schneider. *J. Mol. Biol.* 235, 13-26 (1994).
46. A. Zemla, C. Venclovas, K. Fidelis, and B. Rost. *Proteins* 34, 220-223 (1999).
47. C. A. Orengo, J. E. Bray, T. Hubbard, L. LoConte, and I. Sillitoe. *Proteins Suppl.* 3, 149-170 (1999).
48. T. E. Creighton. *J. Mol. Biol.* 129, 411-431 (1979)
49. K. A. Dill and H. S. Chan. *Nat. Struct. Biol.* 4, 10-19 (1997)
50. H. Roder and W. Colon. *Curr. Opin. Struct. Biol.* 7, 15-28 (1997)
51. R. L. Baldwin. *Science* 295, 1657-1658 (2002)
52. C. M. Dobson. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 356, 133-145 (2001)
53. S. Hayward. *Protein Sci* 10, 2219-2227 (2001)
54. A. M. Gutin, V. I. Abkevich, and E. I. Shakhnovich. *Biochemistry* 34, 3066-3076 (1995)
55. P. X. Qi, T. R. Sosnick, and S. W. Englander. *Nat. Struct. Biol.* 5, 882-884 (1998)
56. T. R. Sosnick, L. Mayne, and S. W. Englander. *Proteins* 24, 413-426 (1996)
57. D. K. Wilkins, S. B. Grimshaw, V. Receveur, C. M. Dobson, J. A. Jones, and L. J. Smith. *Biochemistry* 38, 16424-16431 (1999)
58. C. Tanford. *Adv. Protein. Chem.* 23, 121-282 (1968)
59. C. Tanford, K. Kawahara, and S. Lapanje. *J Am Chem Soc* 89, 729-736 (1967)

Date Received: October 4, 2005

Communicated by the Editor
Jiri Sponer

