Hydrophobic collapse in (*in silico*) protein folding

Michal Brylinski [a,b], Leszek Konieczny [c], Irena Roterman [a,d,*]

[a] *Department of Bioinformatics and Telemedicine, Collegium Medicum,*
Jagiellonian University, Kopernika 17, 31-501 Krakow, Poland
[b] *Faculty of Chemistry, Jagiellonian University, Ingardena 3, 30-060 Krakow, Poland*
[c] *Institute of Medical Biochemistry, Collegium Medicum, Jagiellonian University,*
Kopernika 7, 31 034 Krakow, Poland
[d] *Faculty of Physics, Jagiellonian University, Reymonta 4, 30-060 Krakow, Poland*

Abstract

A model of hydrophobic collapse, which is treated as the driving force for protein folding, is presented. This model is the superposition of three models commonly used in protein structure prediction: (1) 'oil-drop' model introduced by Kauzmann, (2) a lattice model introduced to decrease the number of degrees of freedom for structural changes and (3) a model of the formation of hydrophobic core as a key feature in driving the folding of proteins.

These three models together helped to develop the idea of a fuzzy-oil-drop as a model for an external force field of hydrophobic character mimicking the hydrophobicity-differentiated environment for hydrophobic collapse. All amino acids in the polypeptide interact pair-wise during the folding process (energy minimization procedure) and interact with the external hydrophobic force field defined by a three-dimensional Gaussian function. The value of the Gaussian function usually interpreted as a probability distribution is treated as a normalized hydrophobicity distribution, with its maximum in the center of the ellipsoid and decreasing proportionally with the distance versus the center. The fuzzy-oil-drop is elastic and changes its shape and size during the simulated folding procedure.
© 2006 Elsevier Ltd. All rights reserved.

Keywords: Protein structure prediction; Late-stage folding; Hydrophobic collapse

1. Introduction

Since the classic work by Kauzmann (1959), it has turned out that hydrophobic interactions play a crucial role in forming and stabilizing protein tertiary structure. It is generally accepted that globular proteins consist of a hydrophobic core and a hydrophilic exterior (Klapper, 1971; Klotz, 1970; Kyte and Doolittle, 1982; Meirovitch and Scheraga, 1980a,b, 1981). The way in which the amino acid sequence partitions a protein into its interior and exterior has been described (Chothia, 1975; Rogov and Nekrasov, 2001; Rose and Roy, 1980; Schwartz et al., 2001). Recent experimental discoveries in protein biosynthesis (Ferbitz et al., 2004) as well as recognition of the structure of chaperonins (Braig et al., 1994; Horwich et al., 1999; Houry et al., 1999; Sakikawa et al., 1999; Taneja and Mande, 2001; Wang et al., 1999; Xu et al.,

1997) suggested introduction of an external hydrophobic force field, which stimulates concentration of hydrophobic residues in a central part of the protein molecule with the hydrophilic residues exposed on the protein surface. The model introduced by Kauzmann (1959) suggests that the protein interior is similar to an organic 'oil-drop'. If so, the hydrophilic residues placed mostly on the protein surface should fit well to the interior of the chaperonin molecule. In consequence, the hydrophobic residues seem to be pushed into the central part of the folded protein facilitating creation of the hydrophobic core. Since the hydrophobic effect was suggested to be the dominant driving force in protein folding (Baldwin, 2002; Dill, 1990; Finney et al., 2003; Pace et al., 1996), the presence of an external force field during protein folding seemed necessary. Estimation of the distribution of packing density in general (Kurochkina and Privalov, 1998) and of the hydrophobicity distribution leading to creation of the hydrophobic core in the protein molecule appeared to be the criterion for predicted protein structure in ab initio approaches (Bonneau et al., 2001). Interial ellipsoids for the buried hydrophobic residues,

* Corresponding author. Tel.: +48 12 424 7278; fax: +48 12 422 7764.
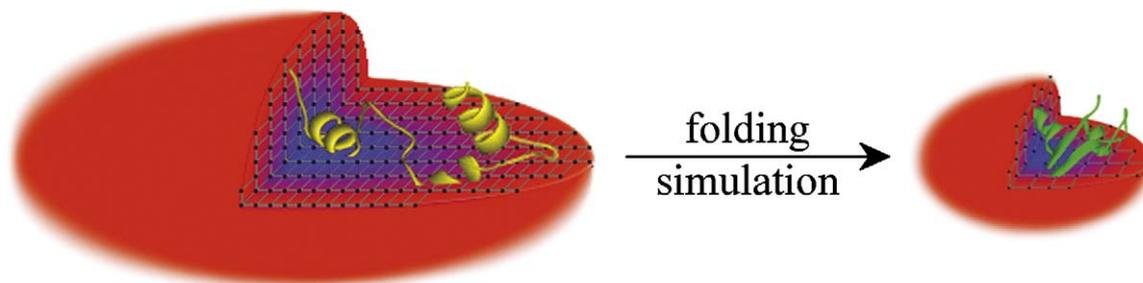 *E-mail address:* myroterm@cyf-kr.edu.pl (I. Roterman).

Fig. 1. Draft view of folding simulation based on the fuzzy-oil-drop model. The spatial hydrophobicity distribution of the fuzzy-oil-drop increases from the hydrophilic exterior layer (red) to the interior hydrophobic core (blue). The folding simulation is driven by the difference between theoretical and observed hydrophobicity at the grid points of the lattice visible inside the drop. The unfolding starting structure of the protein (yellow) is immersed in the drop and folds as the drop size decreases to reach the native form (green).

partitioned according to the proportion (25, 50 and 75%) of all residues localized in particular ellipsoids, were used to estimate the core score and Cα-partitioning score.

On the other hand, vigorously pursued lattice models (Abkevich et al., 1994; Kolinski et al., 2003; Kolinski and Skolnick, 1994; Onuchic et al., 1996; Pande and Rokhsar, 1999; Sali et al., 1994; Skolnick and Kolinski, 1991; Taketomi et al., 1975) allowed a significant decrease of the number of degrees of freedom, assuming that only grids of the lattice can be occupied by particular amino acids (represented by effective atoms). The folding process (*in silico*) based on this model was described in detail (Dobson, 2001), together with the numbers of degrees of freedom at each step of the folding process. Despite the obvious simplifications, lattice models have been shown to resemble real proteins in their collapse transitions, development of secondary and tertiary structure, mutational properties, and folding kinetics (Chan and Dill, 1996; Dill et al., 1995; Pande and Rokhsar, 1999; Yue et al., 1995).

Linking all the above models and experimental observations, a fuzzy-oil-drop model can be built to simulate the external hydrophobic force field creating the environment for the folding molecule (Brylinski et al., 2006b; Konieczny et al., 2006). A tentative scheme of hydrophobic collapse in folding simulation based on the fuzzy-oil-drop is presented in Fig. 1.

The starting structure for hydrophobic collapse simulation can be created according to the early-stage folding model presented elsewhere (Jurkowski et al., 2004a; Roterman, 1995a,b). The model for early-stage folding (*in silico*) was also verified using BPTI (Brylinski et al., 2004a), lysozyme (Jurkowski et al., 2004b) and the α and β chains of hemoglobin (Brylinski et al., 2004b). The limited conformational sub-space and particularly the probability profile of the ϕ, ψ angles in this sub-space allowed generalization and introduction of letter codes for early-stage structure. Summarized in the sequence-to-structure and structure-to-sequence contingency table (Brylinski et al., 2005), they made it possible to predict the early-stage structure of a polypeptide of any amino acid sequence (Brylinski et al., 2004c).

2. Materials and methods

2.1. Theoretical fuzzy-oil-drop

The fuzzy-oil-drop representing the environment for polypeptide folding is described by a three-dimensional Gaussian function. The Gaussian function usually interpreted as a probability distribution is assumed to represent the hydrophobicity distribution. If the jth point described by Cartesian coordinates (x_j, y_j, z_j) belongs to a box covering entirely the molecule and localized with its center at the origin of the coordinate system $(0, 0, 0)$ the theoretical hydrophobicity value $\tilde{H}t_j$ for this point, is calculated as follows:

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

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. $\tilde{H}t_{sum}$ is the sum of theoretical hydrophobicity for all analyzed grid points. In this manner the normalized hydrophobicity value, varying from 0.0 for edge points to its maximum for $(\bar{x}, \bar{y}, \bar{z})$, can be calculated. Each jth grid point (x_j, y_j, z_j) is characterized by the $\tilde{H}t_j$ value, which represents the idealized hydrophobicity density in the fuzzy-oil-drop. The ellipsoid is covered by grid system (5 Å constant step). The $\tilde{H}t_j$ value is calculated for each grid point.

2.2. Observed fuzzy-oil-drop

The observed hydrophobicity distribution within the fuzzy-oil-drop is calculated using the simple sigmoid function previously proposed to quantitative description of the hydrophobic interactions (Levitt, 1976). The jth grid point (same as in theoretical oil-drop) collects hydrophobicity $\tilde{H}o_j$ as follows:

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

where N is the total number of residues in the protein under consideration, $\tilde{H}_i^r$ denotes the hydrophobicity of the ith residue according to the normalized scale of hydrophobicity for amino acids, r_{ij} denotes the separation of the jth grid point and the effective atom of the ith residue, and c denotes the hydrophobic cutoff and has the fixed value of 9.0 Å following the original paper (Levitt, 1976). This means that only residues with $r_{ij} \leq c$ influence the jth point. $\tilde{H}o_{sum}$ is the sum of observed hydrophobicity for all analyzed grid points. Using $1/\tilde{H}o_{sum}$ as a normalizing coefficient the observed hydrophobicity can be compared to the theoretical hydrophobicity described previously.

2.3. The size of the fuzzy-oil-drop

The size of the drop is defined individually for the particular protein molecule. Before the fuzzy-oil-drop size and hydrophobic density distribution can be calculated, the protein molecule should be properly oriented in the space. This orientation is as follows: (1) the geometrical center of the protein molecule is placed at a center of the coordinate system (0, 0, 0); (2) the molecule is oriented with the line linking the two effective atoms representing the longest distance of the molecule lying along the Z-axis; (3) the longest distance between two effective atoms projections on the XY plane is found; (4) the molecule is reoriented according to the line linking the two atoms found in step 3, being parallel to the X-axis; and (5) for this orientation, the largest positive and largest negative Y-coordinates are found. The distances found at points 2, 3 and 5 allow definition of the size of the box in which the whole molecule can be placed. Moreover, the distance between each box edge and the nearest point representing the center of residue interaction should be no shorter than the hydrophobic cutoff used [see Eq. (2)], which has the fixed value 9.0 Å. This condition ensures the disappearance of observed hydrophobicity at the edge points, therefore making the theoretical and observed distributions comparable. The box is filled with an internal three-dimensional grid system. The grid step has a fixed value of 5.0 Å in all simulations.

2.4. Theoretical scale of amino acid hydrophobicity based on fuzzy-oil-drop model

Single-domain proteins selected with the aid of CATH Domain Structure Database (Orengo et al., 1997; Pearl et al., 2000, 2005) were used to calculate the value of the hydrophobicity parameter for each residue. The three-dimensional structures of selected proteins (only that determined by X-ray crystallography) were obtained from Protein Data Bank (Berman et al., 2000). First, for each protein from the database, the set of interactive centers was calculated and the box containing the investigated protein was determined. For each residue's effective atom, the theoretical value of hydrophobicity was assigned according to Eq. (1), where $\tilde{H}t_j$ denotes the theoretical value calculated, and (x_j, y_j, z_j) denotes the Cartesian coordinate of the residue's effective point. In this manner the theoretical value of hydrophobicity was obtained for each residues in the set of proteins. Finally, for each amino acid type the mean value was calculated and the set of 20 obtained hydrophobic param-

eters was normalized. The fuzzy-oil-drop scale of amino acid hydrophobicity was compared with the other theoretically based scales (Eisenberg et al., 1984; Engelman et al., 1986; Hopp and Woods, 1981; Kyte and Doolittle, 1982).

2.5. Starting structure

The ribonuclease A molecule [PDB ID: 5RAT (Tilton et al., 1992)] was taken as a test protein. The early-stage structure of ribonuclease [discussed also in (Jurkowski et al., 2004a)], taken as initial for the hydrophobic collapse presented in this paper, was predicted from the amino acid sequence according to the sequence-to-structure contingency table (Brylinski et al., 2005). The algorithm of early-stage structure prediction is described in detail in (Brylinski et al., 2004c).

2.6. Size change of fuzzy-oil-drop during simulation

The sizes of single-domain proteins present in Protein Data Bank were calculated as box sizes (ellipsoids). The relation of the box size to the number of amino acids in the polypeptide was approximated to a log function. The details of this procedure are presented elsewhere (Brylinski et al., 2006a). The size of the starting structure (early-stage) can also be calculated by the same procedure. According to these relations, the fuzzy-oil-drop was linearly squeezed from early-stage size to the predicted target size in 10 equal steps.

2.7. Optimization procedure

The structure optimization procedure was performed iteratively in 10 main steps. Each main step of optimization consisted of two sub-steps: the first, traditional energy minimization (which controls the folding pathway excluding potential atoms overlaps); and the second, optimization of the hydrophobicity distribution (the main driving force of hydrophobic collapse). Since hydrophobicity distribution optimization takes into account only pair-wise interactions between effective atoms and grid points of the external force field, it may lead to unreal pair-wise contacts between residues. Therefore, the energy minimization procedure is necessary to prevent strangulation of the backbone and to maintain the right folding pathway. The energy minimization sub-step is performed for a full-atom representation of the protein according to ECEPP/3 standards including torsional potential, electrostatic and vdW interactions (Dunfield et al., 1978; Momany et al., 1975; Nemethy et al., 1983, 1992; Sippl et al., 1984). The actual engine of hydrophobic collapse simulation is the second sub-step - hydrophobicity distribution optimization performed for decreased size of fuzzy-oil-drop. This procedure optimizes $\Delta\tilde{H}_{tot}$, which can be interpreted as the difference between theoretical $\tilde{H}t$ and observed $\tilde{H}o$ hydrophobicity over all grid points:

$$\Delta\tilde{H}_{tot} = \sum_{j=1}^{P} (\tilde{H}t_j - \tilde{H}o_j)^2 \tag{3}$$
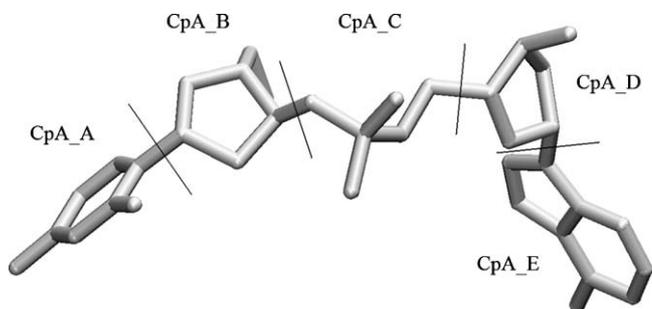
Fig. 2. Five virtual amino acids of CpA molecule: CpA_A, CpA_B, CpA_C, CpA_D and CpA_E. For each virtual amino acids the hydrophobic parameter was assigned independently.

where $\tilde{H}t_j$ and $\tilde{H}o_j$ are the theoretical and observed values of hydrophobicity for the jth grid point, respectively. P denotes the total number of grid points at a particular step of folding simulation.

The algorithm given by Rosenbrock (1960) was applied to optimize $\Delta\tilde{H}_{tot}$ during the simulation.

2.8. Refinement of the final structure

Energy minimization of ribonuclease in its post-hydrophobic collapse form was performed using the AMBER 7.0 program (Pearlman et al., 1995) with the ff99 (Wang et al., 2000) force field. The water environment including the charge-screening effects of salt was represented implicitly by the analytic Generalized Born (GB) model (Bashford and Case, 2000; Onufriev et al., 2002). The surrounding salt concentration was set to physiological conditions (0.2 mol/l). The refinement procedure includes 5000 steps of unconstrained energy minimization; after 2500 steps of steepest descent minimization, the method was switched to the conjugant gradient method.

2.9. Presence of ligand

The results of the folding process simulated according to the procedure presented above indicated the necessary presence of a ligand molecule in the simulation environment during the folding process. The deoxycytidyl-3′,5′-deoxyadenosine (CpA) molecule, a molecule treated as containing five virtual amino acids with their hydrophobic characteristics (Fig. 2), was free to translate into three directions and to rotate around three axes (six degrees of freedom). The position of CpA was dependent on its interaction with the external force field (fuzzy-oil-drop) and on its interaction with the folding polypeptide chain, mutually influencing each other. The presence of CpA molecule during simulated ribonuclease folding required theoretical hydrophobicity parameters to be assigned to its virtual amino acids. The hydrophobicity parameters were calculated using the crystal structure of ribonuclease A complexed with CpA [PDB ID: 1RPG (Zegers et al., 1994)] according to Eq. (1), where $\tilde{H}t_j$ denotes the theoretical value calculated, and (x_j, y_j, z_j) denotes the Cartesian coordinate of the virtual residue's effective point. The partial charges for CpA

have been calculated using InsightII (Molecular Simulations Inc.).

2.10. Structure analysis

Both the early-stage (initial) structure and the one with hydrophobic core simulated were compared with the native structure of this protein according to the following criteria:

1. The total number of non-bonding interactions (NB) assuming a cutoff distance of 12 Å.
2. The accessible surface area (ASA) taking a probe radius of 1.4 Å. ASA was calculated using Surface Racer program (Tsodikov et al., 2002).
3. The radius of gyration (R_g), which was calculated using the following equation (Flory, 1953):

$$R_g = \sqrt{\frac{1}{(N+1)^2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\langle(\vec{r}_i - \vec{r}_j)^2\rangle} \qquad (4)$$

where N is the number of residues, $\vec{r}_i$ and $\vec{r}_j$ are the coordinates of Cα atom of ith and jth residue, respectively,
4. The distances between the geometric center of a molecule and sequential Cα atoms in the polypeptide chain ($D_{center–C\alpha}$).
5. The distances between the geometric center of a ligand molecule and sequential Cα atoms in the polypeptide chain ($D_{ligand–C\alpha}$).
6. RMSD-Cα calculation using the native form of ribonuclease as a reference structure. Structure alignments and RMSD calculations were done using VMD (Humphrey et al., 1996).
7. The distances between Cα atoms of the residues designated and represent the native disulfide bonds.

2.11. 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 ith residue $\Delta\tilde{H}_i$ is calculated as follows:

$$\Delta\tilde{H}_i = \tilde{H}t_i - \tilde{H}o_i \qquad (5)$$

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

Additionally, localization of these irregularities versus the idealized fuzzy-oil-drop allows the determination of function-related areas in protein molecules. The prediction of active site according to fuzzy-oil-drop model can be easily carried out for any protein structure with free prediction server available from http://bioinformatics.cm-uj.krakow.pl/activesite.

3. Results and discussion

3.1. Statistical scale of amino acid hydrophobicity based on the fuzzy-oil-drop model

Hydrophobicity describes solvation in a water solvent. Each residue is characterized by a different degree of hydrophobicity or hydrophilicity. The huge number of protein structures accumulated in databases provides an opportunity to statistically analyze the features of amino acid residues with high accuracy. A new theoretical scale of amino acid hydrophobicity was created as described in Section 2. Table 1 presents the values of the hydrophobicity parameter assigned to each amino acid. The scale based on the fuzzy-oil-drop model was found to be in conformity to the other theoretical-based scales, as expected (Fig. 3). There clearly are some differences between the scales compared, with respect to the placement of specific amino acids. The fuzzy-oil-drop scale places cysteine as the most hydrophobic residue. It is well-known that scales based on analysis of proteins with known 3D structures often define hydrophobic character as the tendency for a residue to be found inside a protein rather than on its surface. In the case of cysteine, because it is involved in disulphide bonds that must necessarily occur inside a globular structure, our results are consistent with some other scales (Janin, 1979; Rose et al., 1985). The strongly hydrophobic nature of cysteine residues in proteins was also reported by (Nagano et al., 1999).

3.2. Early-stage structure prediction

A large body of experiments and theoretical evidence suggests that local structure is frequently encoded in short segments of the protein sequence. Models known in the literature concerning the sequence-to-structure relation discuss protein structure as it appears in the final native form of the protein (Bystroff and Baker, 1998; de Brevern et al., 2000, 2002; Fetrow et al., 1997; Han and Baker, 1996). The model of the ellipse-path-limited conformational sub-space for proteins (Brylinski et al., 2004a,b; Jurkowski et al., 2004a,b) represents an approach to the relation between sequence and structure in the early-stage folding (*in silico*) structural form. The recently published results of molecular dynamics simulation of 3- and 21-alanine polypep-

Table 1
Values of the hydrophobicity parameter ($\tilde{H}_i^r$) assigned to each amino acid according to the fuzzy-oil-drop model

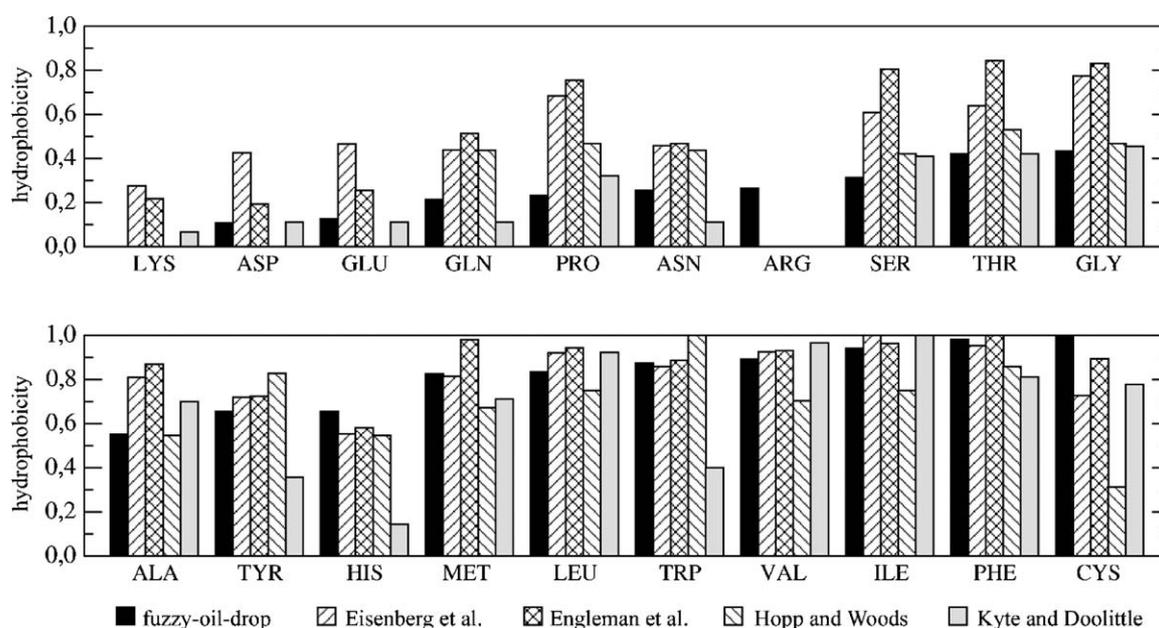| Amino acid	$\tilde{H}_i^r$
LYS	0.000
ASP	0.108
GLU	0.126
GLN	0.215
PRO	0.233
ASN	0.256
ARG	0.265
SER	0.314
THR	0.422
GLY	0.435
ALA	0.552
HIS	0.655
TYR	0.655
MET	0.825
LEU	0.834
TRP	0.874
VAL	0.892
ILE	0.942
PHE	0.982
CYS	1.000

hydrophobicity ↓

Fig. 3. Comparison of hydrophobicity values for amino acids according to different scales. All scales are normalized.
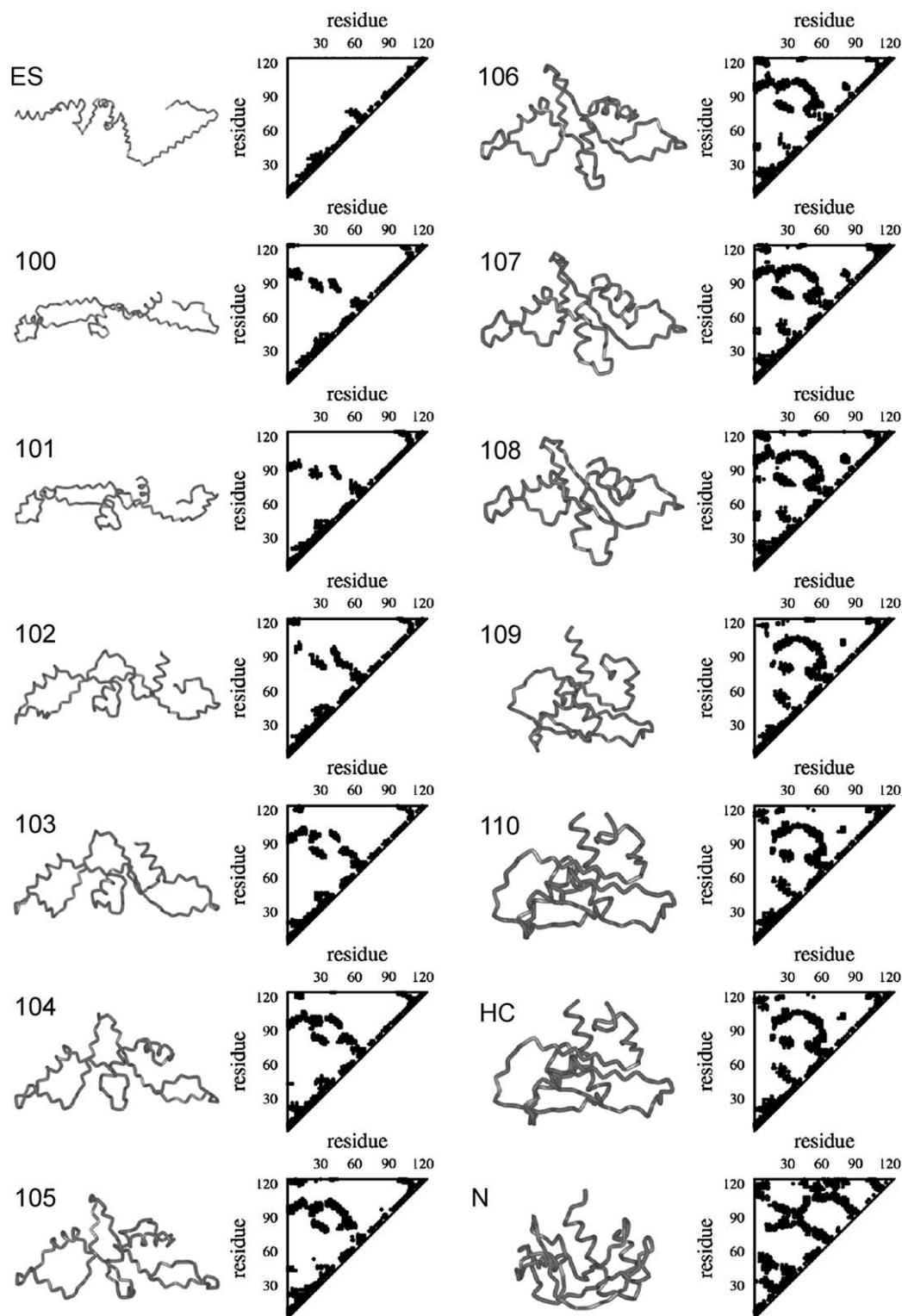
Fig. 4. Snapshots of folding route. ES—early-stage (starting) form of ribonuclease predicted from its amino acid sequence, I00–I10—intermediates obtained after each step of folding simulation, HC—hydrophobic collapse (final), N—native structure obtained from PDB. Maps of non-bonding interactions (NB) were calculated for each structure assuming a 12 Å cutoff.

tide in the temperature range 276–402 K additionally positively verified the ellipse path as the path, along which the helix unfolding is taking place (Gnanakaran and Garcia, 2005; Paschek et al., 2005). The results of this simulation additionally support the reliability, that the ellipse-path limited conformational sub-

space can be used to determine the starting structure for folding process simulation.

Ribonuclease A which is well characterized as a single-domain protein that contains 124 residues with four native disulfide bonds, has played a crucial role as a model system

in studies of protein structure, folding, and enzyme catalysis. The sequence of ribonuclease was used as input for early-stage structure prediction on the basis of a sequence-to-structure contingency table (Brylinski et al., 2005). The Structure Predictability Index (SPI) (Brylinski et al., 2004c) calculated for the sequence was found to be 95.0, which ranks ribonuclease as an easy target for early-stage form prediction. The Q3 (Rost and Sander, 1993), Q7 (Brylinski et al., 2004c) and SOV (Rost et al., 1994; Zemla et al., 1999) parameters calculated versus native structure were found to be 93.4, 94.3 and 86.0, respectively. The high accuracy of early-stage structure prediction for ribonuclease justified the selection of this form as the good starting structure for simulation of hydrophobic core and estimation of its role in protein folding. The details of the early-stage structure prediction method based on a sequence-to-structure contingency table are given in (Brylinski et al., 2004c, 2005). The prediction, including SPI, Q3 and Q7 estimations, can be easily carried out for any protein sequence with free prediction server available from http://bioinformatics.cm-uj.krakow.pl/earlystage.

3.3. Size change of fuzzy-oil-drop during simulation

The size of the fuzzy-oil-drop, calculated as described in Section 2, was found to be $Z = 143.37$ Å, $X = 63.62$ Å and $Y = 59.65$ Å for early-stage, and $Z = 61.77$ Å, $X = 50.23$ Å and $Y = 48.41$ Å for native form of ribonuclease. The procedure presented in (Brylinski et al., 2006a) predicted the target size of the fuzzy-oil-drop to be $Z = 67.23$ Å, $X = 52.90$ Å and $Y = 46.73$ Å. The high accordance of the target (predicted) and native (observed) sizes of the fuzzy-oil-drop ensured a good size-dependent condition for hydrophobic collapse simulation.

3.4. Non-bonding contacts

Non-bonding interactions (NB) present in all discussed structural forms of ribonuclease are shown as contact maps in Fig. 4. The non-bonding contact distributions were similar for final and native structure. Interactions that stabilize the fold are between residues that are well separated along the sequence and therefore away from the diagonal of the plot, where an interaction was defined as occurring when two Cα atoms were within 12 Å of one another. Cotesta et al. (2003) provided a two-dimensional map of average non-bonded interaction energies between residue pairs resulting from molecular dynamics simulation of ribonuclease A and S-protein (obtained by removing the first 20 residues from the native molecule). Most of the strong attractive interactions contributing to fold stabilization in their system were also found in the final form with well-defined hydrophobic core in our system. Moreover, some of them appeared at the beginning of hydrophobic collapse simulation. This may suggest the important role of these interactions in the folding pathway.

A quantitative analysis of the total number of non-bonding contacts (NB) recorded after particular steps of folding is presented in Fig. 5(A). Along with decreasing size of fuzzy-oil-drop during the simulated folding process of ribonuclease, the total
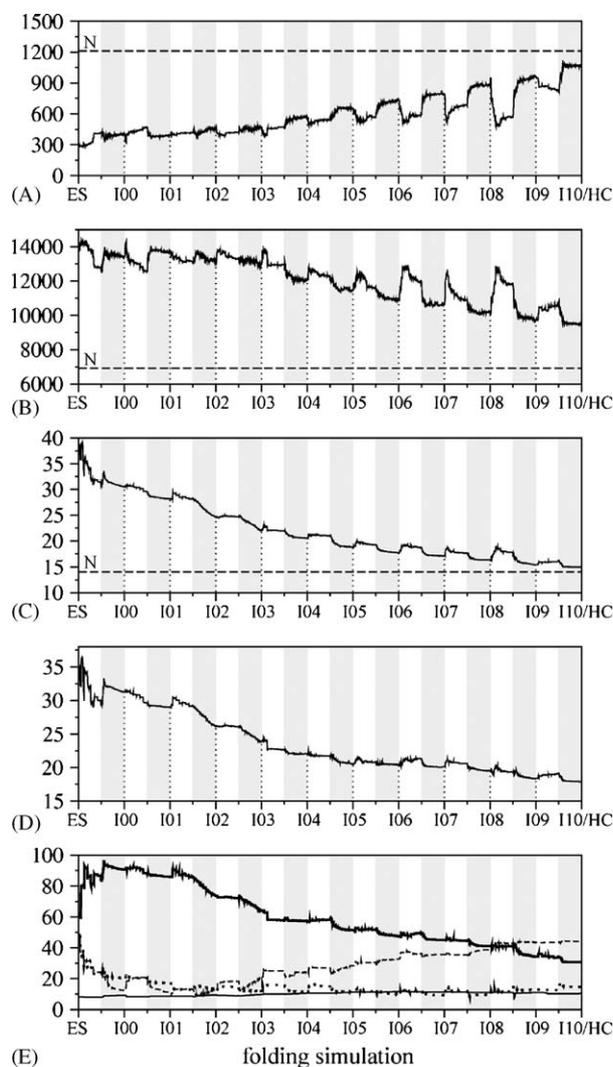

Fig. 5. Profiles of early-stage (ES), hydrophobic core oriented (HC), intermediates of folding simulation (I00–I10) and native form of ribonuclease (N) with respect to: (A) the number of non-bonding contacts (NB), (B) accessible surface area (ASA) [Å²]; (C) radius of gyration (R_g) [Å], (D) RMSD-Cα vs. the native structure [Å], (E) the distances between Cα atoms of the residues C26–C84 (solid), C40–C95 (dashed), C58–C110 (dotted), and C65–C72 (dashed/dotted) which form disulfide bonds in native ribonuclease [Å]. In plots A–C the constant dotted line pinpoints the quantity of a probed parameter (NB, ASA or R_g) calculated for native structure (N). Grey shadowing represent an energy minimization sub-step, while optimization of the hydrophobicity distribution remain white.

number of non-bonding interactions was evenly growing and reached the value slightly below the level characteristic for the native structure at the end of simulation. 30.68% of 1069 non-bonding contacts in the form with hydrophobic core created were also found to be present in the native structure.

3.5. Solvent-exposed surface and radius of gyration analysis

Hydrophobic effects cause nonpolar side-chains to tend to cluster together in the protein interior. The hydrophobic effect can be measured by the solvent-accessible surface area (ASA)

Table 2
Comparison of the radii of gyration (R_g) calculated from the folding simulation with experimental data

Structural form of ribonuclease	R_g [Å]
Native	15.0[a]
	15.4[b]
	14.1[c]
Hydrophobic core oriented[d]	15.0
Heat-denaturated[a]	
Non-reduced	19.3
Reduced	28
Urea-denaturated[b]	
Non-reduced	17.3
Reduced	24
Early-stage of folding[e]	36.6
Random coil[f]	45

[a] Taken from Sosnick and Trewhella (1992).
[b] Taken from Zhou et al. (1998).
[c] Calculated for the crystal structure according to Eq. (4).
[d] Final structure, refined using AMBER7 program.
[e] Initial structure, predicted from amino acid sequence according to the sequence-to-structure contingency table (Brylinski et al., 2004c, 2005).
[f] Calculated by Zhou et al. (1998) from the relationship $6R_g^2 = 130n$, where n is the number of residues, including the correction factor for chain length, as described by Tanford (1968).

of a protein, that is, the part of the complex surface in direct contact with solvent. The radius of gyration (R_g) estimates the characteristic volume of the globular protein and provides quantitative information on its compactness. Step-dependent changes in the accessible surface area and the radius of gyration during the folding step oriented on the hydrophobic core creation of ribonuclease are shown in Fig. 5(B and C), respectively. Both ASA and R_g were decreasing and finally R_g reached the native level, whereas ASA remained above the level characteristic for the native form of ribonuclease. The radius of gyration has been measured by small-angle X-ray scattering for native and denatured states of ribonuclease A (Sosnick and Trewhella, 1992; Zhou et al., 1998). The values of R_g calculated from folding simulation were found to be in good agreement with experimental data (Table 2). While R_g calculated for the early-stage of folding simulation form of ribonuclease is 30–50% greater than of reduced-denatured state reported by experiments, it was found to be smaller than expected for a random coil. At the end of folding simulation R_g achieved exactly the value characteristic for native state suggesting that the compactness of final structure is similar to that of native form of ribonuclease.

3.6. RMSD versus native structure

The changes in RMSD-Cα parameter using the native form of ribonuclease as a reference structure are presented in Fig. 5(D). The unfolded early-stage structure (see Fig. 4) is characterized by large RMSD (32.27 Å). During the folding simulation RMSD was continuously decreasing and reached the value of 18.07 Å for the final structure (hydrophobic core created).

3.7. Disulfide bonds tracing

The distances between Cα atoms involved in disulfide bonds in the native form yield another measure of the overall structure. Fig. 5(E) presents the monitoring of cysteines participating in disulfide bonds during folding simulation of ribonuclease, which has four disulfide bonds (at positions C26–C84, C40–C95, C58–C100, C65–C72). The Cα–Cα distances of appropriate cysteines were measured in sequential steps of the procedure. All the distances (except C65–C72) were found to be long in the early-stage structure, especially that between C58 and C110. The C65–C72 distance remained short and constant during the simulation. The early formation of C65–C72 disulfide bond had already been reported and it was claimed that the C65–C72 loop is located in one of the chain folding initiation sites of ribonuclease (Iwaoka et al., 1998; Rothwarf et al., 1998). Two of the distances significantly decreased: C26–C84 and C58–C110. Interestingly, experimental results revealed that these buried in the native form of ribonuclease disulfide bonds (C26–C84 and C58–C110) contribute more to the stability of the protein than the exposed C40–C95 or C65–C72 bonds (Welker et al., 1999). The distance analysis suggests that three of four disulfide bonds (C65–C72, C40–C95 and C26–C84) are possible to be created in the two initial steps. The next simulation is necessary, during which every cysteines Cα–Cα distance decrease implies disulfide bond creation. It will be expressed by distance constraint. The time dependency of cysteines Cα–Cα distances can be significantly influenced by the presence of such constraints and the 'history' of C58–C100 bond [Fig. 5(E)] may look completely different. The results of such approach will be shown in close future.

3.8. Spatial distribution of Cα atoms versus the geometrical center

The profile of the length of vectors linking the geometrical center with sequential Cα atoms provides insight into the three-dimensional relative displacement versus the native form of the protein (Orengo et al., 1999). Structural similarity may be indicated by overlapping the lines representing the compared structures. Parallel orientation of profiles is interpreted as similarity of structural forms in compared molecules oriented differently in space. The $D_{center–C\alpha}$ vector profiles for early-stage (initial), final one (as resulted after hydrophobic collapse) and native structure presented in Fig. 6 revealed high similarity between the final and native forms of ribonuclease.

3.9. Ligand presence in folding process

The comparison of the native structure of ribonuclease with the one received according to the folding procedure presented in this paper reveals the aim-orientation of the folding process. Fig. 7 shows the distribution of hydrophobicity irregularities $\Delta\tilde{H}$ versus the idealized fuzzy-oil-drop. Since the folding simulation directed by the external force field appeared to produce a structure that was different than expected, the search for the
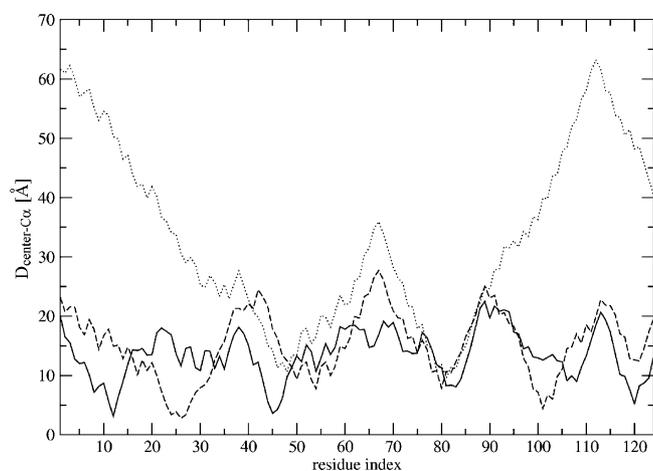
Fig. 6. Profiles of $D_{center-C\alpha}$ vectors for early-stage (dotted line), post-hydrophobic collapse (dashed line) and native (solid line) structure of ribonuclease.

Table 3
Values of hydrophobicity parameter (H_i^r) assigned to each virtual residue of CpA (Fig. 2) according to the fuzzy-oil-drop model

Virtual residue of CpA	$\tilde{H}_i^r$
CpA_A	0.579
CpA_B	0.892
CpA_C	0.797
CpA_D	0.513
CpA_E	0.430

which ensures the creation of a cavity able to bind the ligand molecule in the active site. This observation might explain the source of low quality of predicted structure.

To verify the supposition that the idealized oil-drop can be created in the presence of a foreign molecule, which fitting to the folding polypeptide completes the oil-drop by its own hydrophobicity, additional simulation of ribonuclease in the presence of a CpA molecule was performed. The hydrophobicity parameters assigned to the virtual amino acids of CpA are given in Table 3. The applied procedure was aimed to make the parameterization of CpA molecule as compatible as possible with amino acid parameterization.

The folding process of ribonuclease in the presence of CpA molecule produced better results than the simulation without the
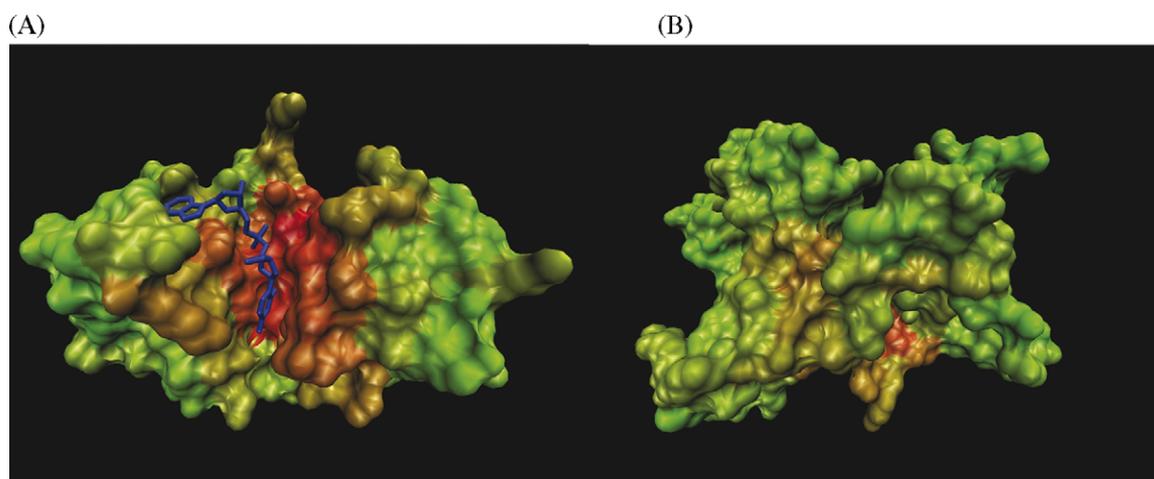
origin of this discrepancy was performed. It turned out that in the native structure of a protein the highest irregularity versus the idealized fuzzy-oil-drop is localized in the active center. It presumably means that the native structure represents the product of aim-oriented folding process. The aim-orientation is understood as directing the folding process by the presence of a ligand,

Fig. 7. The distribution of hydrophobicity irregularities ($\Delta \tilde{H}$) vs. the idealized fuzzy-oil-drop. (A) native structure of ribonuclease complexed with CpA (Zegers et al., 1994) (blue) and (B) post-hydrophobic collapse structural form obtained as the result of 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 visualizes large irregularities ($\Delta \tilde{H}$) in the vicinity of the active site.
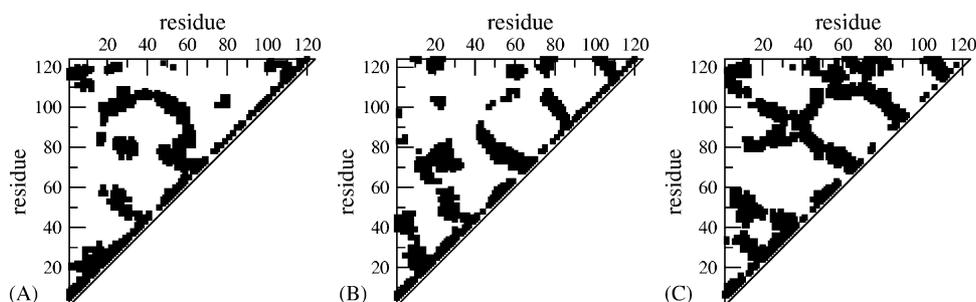

Fig. 8. Non-bonding contact maps for ribonuclease: folded *in silico* in the absence (A) and presence (B) of CpA. For comparison, non-bonding contact maps for native form of ribonuclease are shown in (C).
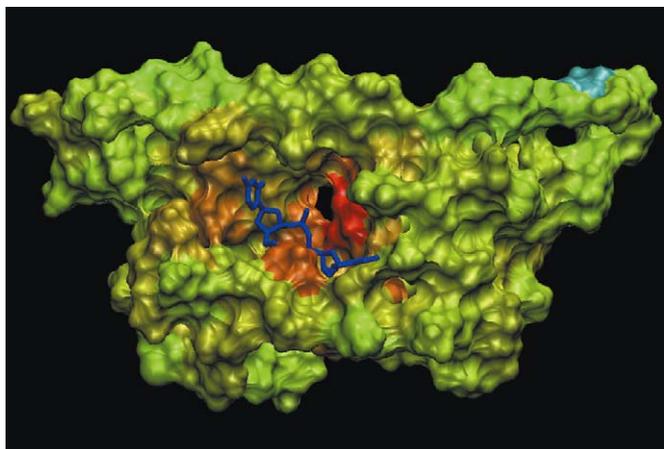
Fig. 9. . The distribution of hydrophobicity irregularities ($\Delta \tilde{H}$) vs. the idealized fuzzy-oil-drop designated for the structure of ribonuclease folded in the active presence of CpA molecule. Color scheme as in Fig. 7.

ligand. The RMSD-Cα parameter calculated using the native form of ribonuclease as a reference structure decreased down to 14.485. Moreover the final contact pattern obtained for the structure of ribonuclease folded in the presence of CpA [Fig. 8(B)] reproduces more of the important features of the native folded structure [Fig. 8(C)] than the contact pattern obtained from the simulation of polypeptide chain alone [Fig. 8(A)]. Although the total number of non-bonding interactions was found to be smaller (NB = 879) compared to the form with hydrophobic core created with no ligand presence during simulation (NB = 1069), the percentage of the reproduced native non-bonding interactions grew to 352 (ligand present) from 328 (ligand absent).

3.10. Expected biological activity

The question arises whether the biological function can be present in molecules received as a result of simulation. So far,

the biological activity might be understood as an aim-oriented discrepancy between ideal and observed hydrophobic oil-drop [Fig. 7(A)]. Fig. 9 shows the distribution of hydrophobicity irregularities $\Delta \tilde{H}$ versus the idealized fuzzy-oil-drop for the structure of ribonuclease folded in the active presence of CpA molecule. The area of great disparity marks the protein's ligation site. Moreover a similar correlation was found between hydrophobicity irregularities $\Delta \tilde{H}$ and the remoteness from a ligand molecule for both native structure and the one folded in the presence of CpA [Fig. 10]. The results suggest that the presence of a natural ligand in the folding process seems to be important to obtain a functionally active protein. The external hydrophobic force field may play an important role directing the folding process toward the specific hydrophobic core creation with the ligand taking part in its construction. One may conclude that the oil-drop model representing the target hydrophobicity distribution may direct the folding process in the proper direction on the condition of active ligand participation in hydrophobic core creation.

4. Conclusions

The fuzzy-oil-drop model seems to work well together with the early-stage model presented elsewhere along with the sequence-to-structure contingency table. Moreover our results seem to be consistent with available experimental observations and folding simulations. The polypeptide structure created according to the limited conformational sub-space was assumed to represent the early-stage of polypeptide chain folding. Ribonuclease A was taken as the test protein. The early-stage structure predicted from the amino acid sequence according to the sequence-to-structure contingency table was an extended low-packed molecule of 1.5–2.5 times bigger than the native form in one dimension. Moreover, most of the secondary structure present in the native structure was already found in the early-stage form. The second step of the folding simulation, which according to many other models represents hydrophobic
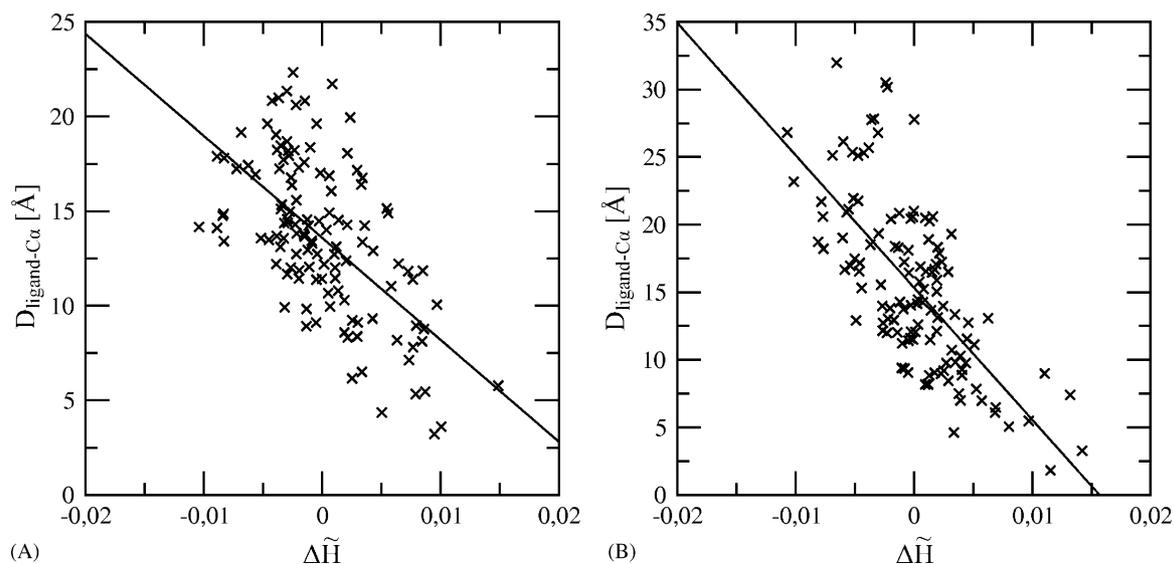

Fig. 10. Correlation between hydrophobicity irregularities ($\Delta \tilde{H}$) and the remoteness from the geometric center of a ligand molecule ($D_{\text{ligand–}C\alpha}$) for native structure of ribonuclease (A) and the structure folded in the presence of CpA (B). The correlation coefficients were found to be −0.60 and −0.69 for A and B, respectively.

collapse, is proposed in this paper. The fuzzy-oil-drop, which in the early stage of folding is represented by a large drop with low hydrophobicity, becomes condensed during the second stage of folding simulation. The change of the three-dimensional Gaussian function (the squeezing process directed by implosion of hydrophobic residues into the central part of the ellipsoid) is accompanied by a simultaneous increase of hydrophobicity density. The detailed analysis of the final structure obtained as a result of folding simulation oriented on hydrophobic core creation revealed distinct similarity to the native one. So far the model presented in this paper was applied to single-domain globular proteins of up to 150 amino acids. The results of folding simulations of lysozyme, BPTI and hypothetical membrane protein–target protein in CASP6 (TA0354_69_121) according to fuzzy-oil-drop model are presented elsewhere (Brylinski et al., 2006b; Konieczny et al., 2006).

The second step of folding (hydrophobic collapse) can be treated continuously in the iteration, allowing the size and shape of the ellipsoid drop to change. In consequence, the folding process goes smoothly, with a well-defined target of hydrophobicity distribution in the protein molecule. The model presented here can be very easily adapted for lattice models, attributing the grid points indicated by the hydrophobicity parameter according to a Gaussian function. The external virtual hydrophobic force field directs the simulated folding process and speeds up creation of the hydrophobic core.

The comparison of the native structure of ribonuclease with the one obtained by the folding procedure presented in this paper reveals the aim-orientation of the folding process. Fig. 7 shows the distribution of hydrophobicity irregularities $\Delta \tilde{H}$ versus the idealized fuzzy-oil-drop. The main difference, very easily recognized, is that in the native structure of the protein the highest irregularity versus the idealized fuzzy-oil-drop is localized in the active center. Using the fuzzy-oil-drop model for protein folding simulation produces a molecule with a regular hydrophobicity distribution, with low hydrophobicity on the surface. Such a molecule would be perfectly soluble without any tendency to interact with any other molecule. This means that the model produces a completely inactive molecule. The results shown in Figs. 9 and 10 precisely demonstrate that aim-orientation is necessary in the folding process. A molecule mimicking the enzyme's substrate should be present to ensure the creation of an active center in the protein molecule. This supposition was verified during the simulation of ribonuclease folding *in silico* (according to the presented model) in the presence of CpA molecule. The results from simulation of folding suggest that the participation of a natural ligand in the folding process *in silico* is important or even essential. The papers reporting the experimental works oriented on folding process describe significant differences between folding kinetics in the absence and presence of a natural ligand. The influence of heme in the folding process of cytochrome was reported in (Garcia et al., 2005).

The introduction of external force field of hydrophobic character in form of three-dimensional Gaussian function (assumed to represent hydrophobicity density distribution) seems to direct the folding process toward the formation of hydrophobic core. The introduction of a natural ligand (or a molecule mimicking it)

into the folding environment seems to direct the folding process toward ligation site creation in a protein molecule and to make it ready to interact with its natural ligand. The results presented in this paper reveal the nonrandom distribution of hydrophobicity discrepancy versus the active site of protein molecule, although the model needs further development.

Acknowledgements

References

Abkevich, V.I., Gutin, A.M., Shakhnovich, E.I., 1994. Specific nucleus as the transition state for protein folding: evidence from the lattice model. Biochemistry 33, 10026–10036.

Baldwin, R.L., 2002. Making a network of hydrophobic clusters. Science 295, 1657–1658.

Bashford, D., Case, D.A., 2000. Generalized born models of macromolecular solvation effects. Annu. Rev. Phys. Chem. 51, 129–152.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein data bank. Nucleic Acids Res. 28, 235–242.

Bonneau, R., Strauss, C.E., Baker, D., 2001. Improving the performance of Rosetta using multiple sequence alignment information and global measures of hydrophobic core formation. Proteins 43, 1–11.

Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D.C., Joachimiak, A., Horwich, A.L., Sigler, P.B., 1994. The crystal structure of the bacterial chaperonin GroEL at 2.8 A. Nature 371, 578–586.

Brylinski, M., Jurkowski, W., Konieczny, L., Roterman, I., 2004a. Limited conformational space for early-stage protein folding simulation. Bioinformatics 20, 199–205.

Brylinski, M., Jurkowski, W., Konieczny, L., Roterman, I., 2004b. Limitation of conformational space for proteins—early stage folding simulation of human α and β hemoglobin chains. TASK Q. 8, 413–422.

Brylinski, M., Konieczny, L., Roterman, I., 2004c. SPI—structure predictability index for protein sequences. In Silico Biol. 5, 0022.

Brylinski, M., Konieczny, L., Czerwonko, P., Jurkowski, W., Roterman, I., 2005. Early-stage folding in proteins (in silico) sequence-to-structure relation. J. Biomed. Biotechnol. 2, 65–79.

Brylinski, M., Konieczny, L., Roterman, I., 2006a. Fuzzy-oil-drop hydrophobic force field—a model to represent late-stage folding (in silico) of lysozyme. J. Biomol. Struct. Dyn. 23, 519–528.

Brylinski, M., Konieczny, L., Roterman, I., 2006 Hydrophobic collapse in late-stage folding (*in silico*) of bovine pancreatic trypsin inhibitor. Biochimie, doi:10.1016/j.biochi.2006.03.008.

Bystroff, C., Baker, D., 1998. Prediction of local structure in proteins using a library of sequence-structure motifs. J. Mol. Biol. 281, 565–577.

Chan, H.S., Dill, K.A., 1996. A simple model of chaperonin-mediated protein folding. Proteins 24, 345–351.

Chothia, C., 1975. Structural invariants in protein folding. Nature 254, 304–308.

Cotesta, S., Tavernelli, I., Di Iorio, E.E., 2003. Dynamics of RNase-A and S-protein: a molecular dynamics simulation of the transition toward a folding intermediate. Biophys. J. 85, 2633–2640.

de Brevern, A.G., Etchebest, C., Hazout, S., 2000. Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks. Proteins 41, 271–287.

de Brevern, A.G., Valadie, H., Hazout, S., Etchebest, C., 2002. Extension of a local backbone description using a structural alphabet: a new approach to the sequence-structure relationship. Protein Sci. 11, 2871–2886.

Dill, K.A., 1990. Dominant forces in protein folding. Biochemistry 29, 7133–7155.

Dill, K.A., Bromberg, S., Yue, K., Fiebig, K.M., Yee, D.P., Thomas, P.D., Chan, H.S., 1995. Principles of protein folding—a perspective from simple exact models. Protein Sci. 4, 561–602.

Dobson, C.M., 2001. The structural basis of protein folding and its links with human disease. Philos. Trans. R. Soc. Lond. B Biol. Sci. 356, 133–145.

Dunfield, L.G., Burgess, A.W., Scheraga, H.A., 1978. Energy parameters in polypeptides. 8. Empirical potential energy algorithm for the conformational analysis of large molecules. J. Phys. Chem. 82, 2609–2616.

Eisenberg, D., Schwarz, E., Komaromy, M., Wall, R., 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J. Mol. Biol. 179, 125–142.

Engelman, D.M., Steitz, T.A., Goldman, A., 1986. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu. Rev. Biophys. Biophys. Chem. 15, 321–353.

Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E., Ban, N., 2004. Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. Nature 431, 590–596.

Fetrow, J.S., Palumbo, M.J., Berg, G., 1997. Patterns, structures, and amino acid frequencies in structural building blocks, a protein secondary structure classification scheme. Proteins 27, 249–271.

Finney, J.L., Bowron, D.T., Daniel, R.M., Timmins, P.A., Roberts, M.A., 2003. Molecular and mesoscale structures in hydrophobically driven aqueous solutions. Biophys. Chem. 105, 391–409.

Flory, P.J., 1953. Principles of Polymer Chemistry. Cornell University Press, Ithaca.

Garcia, P., Bruix, M., Rico, M., Ciofi-Baffoni, S., Banci, L., Ramachandra Shastry, M.C., Roder, H., de Lumley Woodyear, T., Johnson, C.M., Fersht, A.R., Barker, P.D., 2005. Effects of heme on the structure of the denatured state and folding kinetics of cytochrome b562. J. Mol. Biol. 346, 331–344.

Gnanakaran, S., Garcia, A.E., 2005. Helix-coil transition of alanine peptides in water: force field dependence on the folded and unfolded structures. Proteins 59, 773–782.

Han, K.F., Baker, D., 1996. Global properties of the mapping between local amino acid sequence and local structure in proteins. Proc. Natl. Acad. Sci. U. S. A. 93, 5814–5818.

Hopp, T.P., Woods, K.R., 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. U. S. A. 78, 3824–3828.

Horwich, A.L., Weber-Ban, E.U., Finley, D., 1999. Chaperone rings in protein folding and degradation. Proc. Natl. Acad. Sci. U. S. A. 96, 11033–11040.

Houry, W.A., Frishman, D., Eckerskorn, C., Lottspeich, F., Hartl, F.U., 1999. Identification of in vivo substrates of the chaperonin GroEL. Nature 402, 147–154.

Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: visual molecular dynamics. J. Mol. Graph. 14, 33–38, 27–38.

InsightII. San Diego, CA, USA, Molecular Simulations Inc.

Iwaoka, M., Juminaga, D., Scheraga, H.A., 1998. Regeneration of three-disulfide mutants of bovine pancreatic ribonuclease A missing the 65–72 disulfide bond: characterization of a minor folding pathway of ribonuclease A and kinetic roles of Cys65 and Cys72. Biochemistry 37, 4490–4501.

Janin, J., 1979. Surface and inside volumes in globular proteins. Nature 277, 491–492.

Jurkowski, W., Brylinski, M., Konieczny, L., Wiiniowski, Z., Roterman, I., 2004a. Conformational subspace in simulation of early-stage protein folding. Proteins 55, 115–127.

Jurkowski, W., Brylinski, M., Konieczny, L., Roterman, I., 2004b. Lysozyme folded in silico according to the limited conformational sub-space. J. Biomol. Struct. Dyn. 22, 149–158.

Kauzmann, W., 1959. Some factors in the interpretation of protein denaturation. Adv. Protein Chem. 14, 1–63.

Klapper, M.H., 1971. On the nature of the protein interior. Biochim. Biophys. Acta 229, 557–566.

Klotz, I.M., 1970. Comparison of molecular structures of proteins: helix content; distribution of apolar residues. Arch. Biochem. Biophys. 138, 704–706.

Kolinski, A., Gront, D., Pokarowski, P., Skolnick, J., 2003. A simple lattice model that exhibits a protein-like cooperative all-or-none folding transition. Biopolymers 69, 399–405.

Kolinski, A., Skolnick, J., 1994. Monte Carlo simulations of protein folding. I. Lattice model and interaction scheme. Proteins 18, 338–352.

Konieczny, L., Brylinski, M., Roterman, I., 2006. Gauss-function-based model of hydrophobicity density in proteins. In Silico Biol. 6, 0002.

Kurochkina, N., Privalov, G., 1998. Heterogeneity of packing: structural approach. Protein Sci. 7, 897–905.

Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132.

Levitt, M., 1976. A simplified representation of protein conformations for rapid simulation of protein folding. J. Mol. Biol. 104, 59–107.

Meirovitch, H., Scheraga, H.A., 1980a. Empirical studies of hydrophobicity. 1. Effect of protein size on the hydrophobic behavior of amino acids. Macromolecules 13, 1398–1405.

Meirovitch, H., Scheraga, H.A., 1980b. Empirical studies of hydrophobicity. 2. Distribution of the hydrophobic, hydrophilic, neutral, and ambivalent amino acids in the interior and exterior layers of native proteins. Macromolecules 13, 1406–1414.

Meirovitch, H., Scheraga, H.A., 1981. Empirical studies of hydrophobicity. 3. Radial distribution of clusters of hydrophobic and hydrophilic amino acids. Macromolecules 14, 340–345.

Momany, F.A., McGuire, R.F., Burgess, A.W., Scheraga, H.A., 1975. Energy parameters in polypeptides. VII. Geometric parameters, partial charges, non-bonded interactions, hydrogen bond interactions and intrensic torsional potentials for naturally ocurring amino acids. J. Phys. Chem. 79, 2361–2381.

Nagano, N., Ota, M., Nishikawa, K., 1999. Strong hydrophobic nature of cysteine residues in proteins. FEBS Lett. 458, 69–71.

Nemethy, G., Gibson, K.D., Palmer, K.A., Yoon, C.N., Paterlini, G., Zagari, A., Rumsey, S., Scheraga, H.A., 1992. Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. J. Phys. Chem. 96, 6472–6484.

Nemethy, G., Pottle, M.S., Scheraga, H.A., 1983. Energy paramters in polypeptides. 9. Updating of geometrical parameters, non-bonding interactions and hydrogen bonding interactions for naturally occuring amino acids. J. Phys. Chem. 87, 1883–1887.

Onuchic, J.N., Socci, N.D., Luthey-Schulten, Z., Wolynes, P.G., 1996. Protein folding funnels: the nature of the transition state ensemble. Fold Des. 1, 441–450.

Onufriev, A., Case, D.A., Bashford, D., 2002. Effective born radii in the generalized born approximation: the importance of being perfect. J. Comput. Chem. 23, 1297–1304.

Orengo, C.A., Bray, J.E., Hubbard, T., LoConte, L., Sillitoe, I., 1999. Analysis and assessment of ab initio three-dimensional prediction, secondary structure, and contacts prediction. Proteins Suppl. 3, 149–170.

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., Thornton, J.M., 1997. CATH-a hierarchic classification of protein domain structures. Structure 5, 1093–1108.

Pace, C.N., Shirley, B.A., McNutt, M., Gajiwala, K., 1996. Forces contributing to the conformational stability of proteins. FASEB J. 10, 75–83.

Pande, V.S., Rokhsar, D.S., 1999. Folding pathway of a lattice model for proteins. Proc. Natl. Acad. Sci. U. S. A. 96, 1273–1278.

Paschek, D., Gnanakaran, S., Garcia, A.E., 2005. Simulations of the pressure and temperature unfolding of an alpha-helical peptide. Proc. Natl. Acad. Sci. U. S. A. 102, 6765–6770.

Pearl, F., Todd, A., Sillitoe, I., Dibley, M., Redfern, O., Lewis, T., Bennett, C., Marsden, R., Grant, A., Lee, D., Akpor, A., Maibaum, M., Harrison, A., Dallman, T., Reeves, G., Diboun, I., Addou, S., Lise, S., Johnston, C., Sillero, A., Thornton, J., Orengo, C., 2005. The CATH domain structure database and related resources gene3D and DHS provide comprehensive domain family information for genome analysis. Nucleic Acids Res. 33, D247–D251.

Pearl, F.M., Lee, D., Bray, J.E., Sillitoe, I., Todd, A.E., Harrison, A.P., Thornton, J.M., Orengo, C.A., 2000. Assigning genomic sequences to CATH. Nucleic Acids Res. 28, 277–282.

Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.R., Cheatham, T.E., DeBolt, I.S., Ferguson, D., Seibel, G., Kollman, P.A., 1995. AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules. Comp. Phys. Commun. 91, 1–41.

Rogov, S.I., Nekrasov, A.N., 2001. A numerical measure of amino acid residues similarity based on the analysis of their surroundings in natural protein sequences. Protein Eng. 14, 459–463.

Rose, G.D., Geselowitz, A.R., Lesser, G.J., Lee, R.H., Zehfus, M.H., 1985. Hydrophobicity of amino acid residues in globular proteins. Science 229, 834–838.

Rose, G.D., Roy, S., 1980. Hydrophobic basis of packing in globular proteins. Proc. Natl. Acad. Sci. U. S. A. 77, 4643–4647.

Rosenbrock, H.H., 1960. An automatic method for finding the greatest or least value of a function. Comput. J. 3, 175–184.

Rost, B., Sander, C., 1993. Prediction of protein secondary structure at better than 70% accuracy. J. Mol. Biol. 232, 584–599.

Rost, B., Sander, C., Schneider, R., 1994. Redefining the goals of protein secondary structure prediction. J. Mol. Biol. 235, 13–26.

Roterman, I., 1995a. The geometrical analysis of peptide backbone structure and its local deformations. Biochimie 77, 204–216.

Roterman, I., 1995b. Modelling the optimal simulation path in the peptide chain folding-studies based on geometry of alanine heptapeptide. J. Theor. Biol. 177, 283–288.

Rothwarf, D.M., Li, Y.J., Scheraga, H.A., 1998. Regeneration of bovine pancreatic ribonuclease A: detailed kinetic analysis of two independent folding pathways. Biochemistry 37, 3767–3776.

Sakikawa, C., Taguchi, H., Makino, Y., Yoshida, M., 1999. On the maximum size of proteins to stay and fold in the cavity of GroEL underneath GroES. J. Biol. Chem. 274, 21251–21256.

Sali, A., Shakhnovich, E., Karplus, M., 1994. Kinetics of protein folding. A lattice model study of the requirements for folding to the native state. J. Mol. Biol. 235, 1614–1636.

Schwartz, R., Istrail, S., King, J., 2001. Frequencies of amino acid strings in globular protein sequences indicate suppression of blocks of consecutive hydrophobic residues. Protein Sci. 10, 1023–1031.

Sippl, M.J., Nemethy, G., Scheraga, H.A., 1984. Intermolecular potentials from crystal data. 6. Determination of empirical potentials for O—H:O=C hydrogen bonds from packing configurations. J. Phys. Chem. 88, 6231–6233.

Skolnick, J., Kolinski, A., 1991. Dynamic Monte Carlo simulations of a new lattice model of globular protein folding, structure and dynamics. J. Mol. Biol. 221, 499–531.

Sosnick, T.R., Trewhella, J., 1992. Denatured states of ribonuclease A have compact dimensions and residual secondary structure. Biochemistry 31, 8329–8335.

Taketomi, H., Ueda, Y., Go, N., 1975. Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. Int. J. Pept. Protein Res. 7, 445–459.

Taneja, B., Mande, S.C., 2001. Metal ions modulate the plastic nature of Mycobacterium tuberculosis chaperonin-10. Protein Eng. 14, 391–395.

Tanford, C., 1968. Protein denaturation. Adv. Protein Chem. 23, 121–282.

Tilton Jr., R.F., Dewan, J.C., Petsko, G.A., 1992. Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-A at nine different temperatures from 98 to 320 K. Biochemistry 31, 2469–2481.

Tsodikov, O.V., Record Jr., M.T., Sergeev, Y.V., 2002. Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. J. Comput. Chem. 23, 600–609.

Wang, J., Cieplak, P., Kollman, P.A., 2000. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J. Comput. Chem. 21, 1049–1074.

Wang, Z., Feng, H., Landry, S.J., Maxwell, J., Gierasch, L.M., 1999. Basis of substrate binding by the chaperonin GroEL. Biochemistry 38, 12537–12546.

Welker, E., Narayan, M., Volles, M.J., Scheraga, H.A., 1999. Two new structured intermediates in yhe oxidative folding of RNase A. FEBS Lett. 460, 477–479.

Xu, Z., Horwich, A.L., Sigler, P.B., 1997. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. Nature 388, 741–750.

Yue, K., Fiebig, K.M., Thomas, P.D., Chan, H.S., Shakhnovich, E.I., Dill, K.A., 1995. A test of lattice protein folding algorithms. Proc. Natl. Acad. Sci. U. S. A. 92, 325–329.

Zegers, I., Maes, D., Dao-Thi, M.H., Poortmans, F., Palmer, R., Wyns, L., 1994. The structures of RNase A complexed with 3′-CMP and d(CpA): active site conformation and conserved water molecules. Protein Sci. 3, 2322–2339.

Zemla, A., Venclovas, C., Fidelis, K., Rost, B., 1999. A modified definition of Sov, a segment-based measure for protein secondary structure prediction assessment. Proteins 34, 220–223.

Zhou, J.M., Fan, Y.X., Kihara, H., Kimura, K., Amemiya, Y., 1998. The compactness of ribonuclease A and reduced ribonuclease A. FEBS Lett. 430, 275–277.