

Hydrophobic collapse in late-stage folding (in silico) of bovine pancreatic trypsin inhibitor

Michal Brylinski^{a,b}, Leszek Konieczny^c, Irena Roterman^{a,d,*}

^aDepartment of Bioinformatics and Telemedicine, Collegium Medicum, Jagiellonian University, Kopernika 17, 31-501 Cracow, Poland

^bFaculty of Chemistry, Jagiellonian University, Ingardena 3, 30-060 Cracow, Poland

^cInstitute of Biochemistry, Collegium Medicum, Jagiellonian University, Kopernika 7, 31-501 Cracow, Poland

^dFaculty of Physics, Jagiellonian University, Reymonta 4, 30-060 Krakow, Poland

Received 7 December 2005; accepted 23 March 2006

Available online 07 April 2006

Abstract

Hydrophobic collapse is commonly considered as a process of significance for protein folding. Many models have been applied for description of this event. A model introducing an external force field mimicking the hydrophobic environment and simultaneously the driving force for the folding process is presented in this paper. Bovine pancreatic trypsin inhibitor (BPTI) was taken as a test protein. An early-stage folding (in silico) model presented elsewhere was used to create the starting structure for hydrophobic collapse. The resulting structure was energy-refined using molecular dynamics simulation in an explicit solvent. The similarity versus the crystal structure of BPTI is estimated using visual analysis, residue–residue contacts, φ , ψ angle distributions, RMSD, accessible solvent area, radii of gyration and hydrodynamic radii. A program allowing creation of early-stage folding structural forms to be created for any protein is available from <http://bioinformatics.cm-uj.krakow.pl/earlystage>. The program for late-stage folding simulation is available on request.

© 2006 Elsevier SAS. All rights reserved.

Keywords: Protein folding; Hydrophobic collapse; Late-stage folding; Hydrophobicity

1. Introduction

Protein folding is a collective self-organization process, which proceeds via obligatory folding intermediates [1]. The classical models of protein folding, the hydrophobic collapse [2,3], the framework model [4] and the diffusion–collision model [5,6] claim that proteins acquire compactness in a distinctly dynamic process separated from the secondary structure formation. The framework and diffusion–collision models focus on the formation of secondary structural elements followed by their assembly, whereas the hydrophobic collapse emphasizes the formation of tertiary structure accompanying secondary structure formation. However, it is difficult to clearly differentiate between these folding scenarios by experimental observation of natural proteins because of the short lifetimes of the folding intermediates [7,8]. Theoretical investigations

suggest that main-chain collapse and secondary structure formation are mostly concerted [9]. More recently, a theoretical model based on a statistical landscape analysis has been proposed [10]. The landscape perspective readily explains the process of reaching a global minimum in free energy and doing so quickly by multiple folding routes on funnel-like energy landscapes [11]. The free energy surface may be constructed by using several different order parameters such as topological contact, radius of gyration, etc. The landscape of proteins is not globally flat, but has a preferred direction of flow toward the native fold, commonly represented as a rugged funnel, the so-called folding funnel [10,12].

The classical ‘oil-drop’ model introduced by Kauzmann [13] represents well the phenomena that accompany the type of protein folding called hydrophobic collapse. According to experimental observations, the second step of folding is driven mainly by hydrophobic interactions [14–17]. It has long been observed that residues with hydrophobic side-chains tend to segregate into the interior of a globular protein, thus constituting a hydrophobic core in which they interact with each other

* Corresponding author.

E-mail address: myroterm@cyf-kr.edu.pl (I. Roterman).

rather than with water, whereas polar and charged residues remain exposed to the solvent [18–23]. The spatial distribution of amino acid hydrophobicity is widely used for validating predicted protein structures [24–27], as well as for identifying the nucleation sites initiating protein folding [28–30]. Calculations incorporating a set of nested ellipsoids, a scaling of the values of amino acid hydrophobicity, and hydrophobic moments have been used to probe this spatial distribution in detail [31–33].

Simplified representations of polypeptide chains mimicking the most pronounced features of globular proteins were developed to improve the sampling of conformational space. Lattice models limit the continuous space to a discrete one, where only grid points are treated as potential positions of atoms in space [10,34–38]. Models for heteropolymer folding on a lattice limit significantly the number of degrees of freedom in conformational space exploration, and therefore have been useful for examining principles of protein folding and addressing problems of conformational change that are too large to be treated by more microscopic models [39–42]. Although lattice models often omit features critical for understanding protein function, they qualitatively mimic way the most pronounced features of globular proteins [43,44].

These two models put together allow the creation of a *fuzzy-oil-drop* model as a tool for hydrophobic collapse simulation. In contrast to the hydrophobic core, commonly reached as a consequence of residue–residue interaction leading to concentration of highly hydrophobic residues in the center of the molecule, our model introduces an external hydrophobic force field in the form of a three-dimensional Gauss function. A draft view of the *fuzzy-oil-drop* model is presented in Fig. 1. The polypeptide chain folds according to an optimization procedure minimizing the difference between the virtual hydrophobicity value at each grid point of the external force field and the inter-residual hydrophobic interaction value attributed to the particular grid point using the function proposed by Levitt [45]. The

standardized Gaussian function parameters control the folding, keeping the mean value fixed in the center of the ellipsoid and decreasing the size of the drop during the folding process (in silico), which increases the packing.

This paper presents the *fuzzy-oil-drop* model and its application to bovine pancreatic trypsin inhibitor (BPTI) folding.

2. Materials and methods

2.1. Early-stage folding simulation

Early-stage folding (in silico) has been presented elsewhere [46–48] and applied to folding simulation of BPTI [49], ribonuclease [46], lysozyme [50] and hemoglobin α and β chains [51]. The main idea for early-stage model is based on the commonly accepted assumption that the initial steps of structure formation are mainly backbone conformation-dependent [40, 47,52].

When all available proteins were transformed to their early-stage folding structural forms (moving the ϕ , ψ angles toward ellipse according to the shortest distance criterion), seven probability maxima could be distinguished in the ϕ , ψ distribution along the path assumed to represent the limited conformational sub-space [53]. These well distinguished maxima allowed the introduction of letter codes for structural motifs (particular fragments on the ellipse-path). On the basis of this observation, a sequence-to-structure contingency table was created [53], expressing the relation between these two characteristics. The sequence expressed by four letters (traditional one-letter code for amino acids) and structure expressed by four-letter codes (expressing the corresponding probability maxima on the ellipse-path) can be ordered in the form of a contingency table. This table, each cell of which expresses the probability of a particular tetrapeptide (sequence) to represent a particular structural motif (early-stage folding form), can be further used to assign the structure for a given amino acid sequence [54]. Analysis of

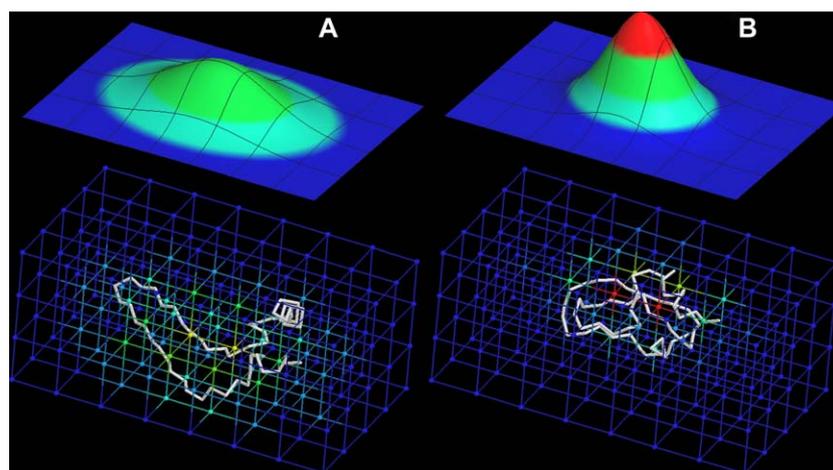


Fig. 1. Diagram showing the relation between the spatial hydrophobicity distribution in the *fuzzy-oil-drop* and the structure of the folding chain. The initial early-stage form of BPTI (A) is immersed in the drop and folds to reach the target native form (B). Top panel: 3D plots displaying the virtual hydrophobicity distribution in a plane horizontally crossing the *fuzzy-oil-drop* by halves. Bottom panel: virtual hydrophobicity attributed to the grid points of an external lattice. The virtual hydrophobicity is colored according to a rainbow scale: blue (the hydrophilic exterior)–green (the middle layer)–red (the interior hydrophobic core). During the simulation a stable hydrophobic core (red) is expected to be constituted.

the general characteristics of the contingency table [53] and the introduction of the Structure Predictability Index (SPI) [54] were intended to create a library of early-stage folding structures for any amino acid sequence. The early-stage structure of BPTI was predicted from the amino acid sequence according to the sequence-to-structure contingency table and subjected to late-stage folding simulation presented in this paper as the initial structure.

2.2. Simplified geometry

Each residue is generally characterized by a different degree of hydrophobicity or hydrophilicity. A first step towards the simplification of the protein structure is to combine groups of atoms belonging to particular side-chain into single effective atom. An effective atom is placed in the geometrical center of the side-chain defined as the middle of the distance between its two the most remote atoms. In the presented simplified representation, the side-chains are rigid and stick out from the backbone. As in conventional representations of polypeptides, the backbone degrees of freedom are the φ and ψ torsion angles about the N–C α and C α –C bonds, respectively. The peptide bond ω dihedral angles are kept fixed at the energy-minimized values.

2.3. Initial orientation of a protein

The *fuzzy-oil-drop* model requires the protein molecule to be properly oriented in the space. The orientation is carried out as follows:

- the geometrical center of the protein molecule is placed at a center of the coordinate system (0, 0, 0);
- the longest distance between two effective atoms was taken as the D_Z measure;
- the longest distance between two projections (on XY -plane) of effective atoms was taken as the D_X measure,
- the difference between the highest and lowest values of Y was taken as the measure of the D_Y ;
- the D_X , D_Y and D_Z was extended by the cutoff distance for hydrophobic interaction (for reasons given in Appendix), which has the fixed value of 9.0 Å;
- the elongated D_X , D_Y and D_Z (according to point 5) are used to determine the box covering the protein molecule entirely. The box of $D_Z \times D_X \times D_Y$ and size determines the values of standard deviations for each one-dimensional representation ($\sigma_Z, \sigma_X, \sigma_Y$).

2.4. Grid system creation

The box is filled with an internal three-dimensional grid similar to the ones used in lattice models [10,34–38]. A constant 5 Å grid size was chosen to split the difference between computational time and accuracy.

2.5. The size of fuzzy-oil-drop

The control of *fuzzy-oil-drop* size is critical for the folding process in our model. Late-stage folding simulation proceeds with the downsizing of *fuzzy-oil-drop* to reach the volume characteristic for the native state. The degree of expected drop compression from early-stage to native conformational state was estimated quantitatively and presented elsewhere [55]. A method to predict the molecular dimension of native state from a polypeptide chain length was proposed to hold the ab initio status. Some essential details are given in Appendix. According to these relations, the *fuzzy-oil-drop* was linearly squeezed from early-stage size to the predicted (expected to be native) size in 10 equal steps.

2.6. Hydrophobicity density calculation

Each grid point of the internal lattice is described by two types of hydrophobicity:

- the theoretical (expected) hydrophobicity $\tilde{H}t_j$ was calculated according to the three-dimensional Gauss function for Cartesian coordinates and standard deviations (the formula is given in Appendix);
- the observed (empirical) hydrophobicity $\tilde{H}o_j$, which expresses the real hydrophobicity derived from nearby amino acids located within the radius of 9 Å (the formula is given in Appendix);

Both types of hydrophobicity are independently standardized (the sum of $\tilde{H}t$ as well as $\tilde{H}o$ for all grid points is equal to 1.0).

2.7. Hydrophobicity density optimization

Each step of *fuzzy-oil-drop* downsizing is followed by structural changes. The structural changes lowering hydrophobicity differences $\Delta\tilde{H}_{tot}$ were accepted. The $\Delta\tilde{H}_{tot}$ parameter can be interpreted as the difference between theoretical $\tilde{H}t$ and observed $\tilde{H}o$ hydrophobicity all over the grid points:

$$\Delta\tilde{H}_{tot} = \sum_{j=1}^P (\tilde{H}t_j - \tilde{H}o_j)^2$$

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

The algorithm given by Rosenbrock [56] was applied to optimize $\Delta\tilde{H}$ during the simulation.

2.8. Late-stage folding simulation

Late-stage folding simulation is performed iteratively in 10 main steps. Each main step consists of two sub-steps: traditional energy minimization (which controls the folding path-

way excluding possible atoms overlaps) followed by the optimization of the hydrophobicity density for decreasing drop size (the main driving force of late-stage folding). Since hydrophobicity density 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 [57–61]. The hydrophobicity density optimization sub-step is performed for the simplified geometry of a protein. Nevertheless, the effective atoms are calculated for each residue au courant to avoid the ambiguity of the conversion from simplified to full representation of a protein during the sub-step switch.

2.9. Refinement of the final structure

The refinement of the late-stage structure of BPTI was performed in explicit solvent with periodic boundaries using the AMBER 7.0 program [62] with the ff99 [63] force field. Six chloride ions were added at positions of high positive electric potential around protein molecule in order to neutralize it. A rectangular box of pre-equilibrated water (WATBOX216) was added to the system resulted in the total of 2935 water molecules to form a box of $54.78 \times 53.03 \times 46.88$ Å. Long range non-bonded interactions were truncated by using a 10 Å cutoff (electrostatic and VdW). Bond constraints were imposed on all bonds involving hydrogen atoms via the SHAKE algorithm [64]. The time-step length was 2 fs and the non-bonded pair list was updated every 10 step during the simulation. The coordinates were saved at every 1 ps. The refinement procedure for solvated system consisted of following steps:

- Initial energy minimization procedure for solvated protein consisted of a two-stage approach. In the first stage the protein was kept fixed while the positions of the water and ions were minimized in 2000 steps. After 1000 steps of the steepest descent minimization the method was switched to the conjugant gradient method. The protein atoms were constrained to their original positions with a force constant $k = 100.0$ [kcal/mol per Å²]. Then in the second stage the entire system was minimized in another 2000 steps.
- Twenty picoseconds of constant volume equilibration, during which the temperature of the system were gradually raised from 0 to 300 K. The protein molecule was constrained to its energy-minimized structure by a weaker potential of 10.0 [kcal/mol per Å²]. At every 1000 steps the translational and rotational motion was removed.
- Hundred picoseconds of unconstrained MD simulation at constant temperature and pressure.
- Production run—1 ns of unconstrained MD simulation at constant temperature and volume.
- Final unconstrained energy minimization consisted of 1000 steps of steepest descent minimization, followed by 1000 steps of conjugant gradient minimization.

2.10. Structure comparison

The early-stage and late-stage structures were compared with the native structure of BPTI to estimate the approach toward the proper structure of this protein. The accordance of the predicted structures with the native one was estimated using the following methods:

- Visual analysis of 3D models.
- Visually judging the similarity of φ , ψ angles distribution over the whole Ramachandran map.
- Comparison of the profiles of the vector length linking the geometric center of the molecule with sequential C α atoms ($D_{\text{center-C}\alpha}$).
- Comparison of the residue–residue contacts (R–R) in the model with the residue–residue contacts in native structure, allowing visualization of the packing pattern.
- Calculating the accessible surface area (ASA) taking a probe radius of 1.4 Å, using the Surface Racer program [65].
- Calculating the radii of gyration (R_g) and hydrodynamic radii (R_h) using the program HYDROPRO [66].
- Calculating RMSD-C α using the native form of BPTI as a reference structure. The structure alignments and RMSD calculations were done using VMD [67].

3. Results and discussion

3.1. Early-stage structure prediction

Different protein sequences can adopt approximately the same 3D structure, and the patterns of sequence conservation can be used for protein structure prediction. While most models concerning the sequence-to-structure relation discuss the structure of proteins as it appears in the final native form of the protein [68–72], the model of an ellipse-path limited conformational sub-space for proteins [46,49–51] represents an approach for 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 polypeptide in the temperature range 276–402 K additionally positively verified the ellipse-path as the path, along which the helix unfolding is taking place [73,74]. 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.

BPTI has long served as an important model system for studies of the protein folding process. The amino acid sequence of BPTI was used as the query input to predict the early-stage form of this protein according to a procedure described previously in [54]. The SPI calculated for the sequence was found to be 95.0, which ranks BPTI as a relative easy target in respect

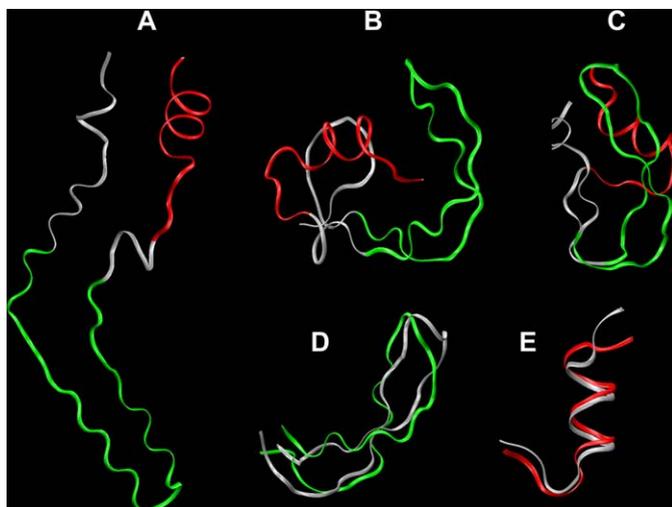


Fig. 2. Ribbon representations of all discussed structural forms of BPTI: early-stage – initial (A), late-stage – final (B) and native – reference structure (C). The main structural elements, the central β -hairpin (13–38) and C-terminal α -helix (44–58), are colored green and red, respectively. Superimposed fragments of β -hairpin (D) and α -helix (E) taken from late-stage (colored) and native (gray) structure.

to the difficulty of early-stage form prediction. Q3 [75], Q7 [54] and SOV [76] parameters calculated versus native structure were found to be 91.1, 87.5 and 89.7, respectively. The high accuracy of early-stage structure prediction justified the selection of this form as a good starting structure for late-stage folding simulation. A 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>.

3.2. Size change of fuzzy-oil-drop during the simulation

The size of the *fuzzy-oil-drop* ($D_z \times D_x \times D_y$) calculated as described in Section 2 was found to be $68.728 \times 47.627 \times 36.459$ and $55.698 \times 41.114 \times 37.683$ Å for the early-stage and native form of BPTI, respectively. The procedure presented in [55] predicted the target size of the *fuzzy-oil-drop* to be $50.206 \times 42.675 \times 39.663$ Å. A fairly high accordance was found for the native and target sizes of the *fuzzy-oil-drop*.

3.3. Visual analysis of 3D models

Fig. 2 visualizes the structural similarities in different structural forms of BPTI. The RMSD-C α values calculated for both early-stage and late-stage models using the native form of BPTI as a reference structure are 13.60 and 10.22, respectively

(Table 1). The colors distinguish the fragments with high structural similarity in the starting early-stage (Fig. 2A), final late-stage (Fig. 2B) and native form (Fig. 2C). The central antiparallel β -sheet (residues 13–38) is colored green, while the C-terminal α -helix (residues 44–58) is red. The RMSD calculated for the distinguished fragments in the late-stage (final) versus native form was found to be 4.23 Å for β -sheet and 2.50 Å for α -helix. Both fragments were found already in the early-stage form of BPTI, suggesting a strong relation between sequence and local structure. β -Sheet folding of fragment (16–36) of BPTI as predicted by Monte Carlo-simulated annealing [77] suggests that the tendency for the peptide segments to form extended β -strands is strong for those with residues 18–24, and moderate for those with residues 30–35. It is noteworthy that right-handed twists of the central β -sheet appeared in the final late-stage form of BPTI. A strong tendency towards right-handed twisting of the β -sheet was also observed in conformational energy minimization [78] and simulated annealing folding studies [77]. Our results support the idea proposed by Daggett and Levitt [79] that stable extended β -sheet may appear early in folding to act as a framework onto which further docking of structure can occur. When all or more segments have been added to the structure, it can then bend and twist to better optimize tertiary interactions.

3.4. φ , ψ Angle changes

Backbone dihedral angles in proteins of known structure lie well inside the allowed regions of a φ , ψ map, to the extent that Ramachandran plot is used routinely to assess the quality of protein structures. The distributions of φ , ψ dihedral angles calculated for all discussed structural forms of BPTI over the whole Ramachandran map are shown in Fig. 3. The plots show the proper migration of central β -sheet residues (green points) in the direction of C7eq during the late-stage folding simulation, while residues of the C-terminal α -helix (red points) remained mostly within the right-handed helix area of the Ramachandran map. The preference of the polypeptide backbone to adopt certain φ , ψ dihedral angles seems to be the consequences of pair-wise interactions between effective atoms and grid points of the external hydrophobic force field affecting the backbone conformation, further supported by the alternating energy minimization procedures carried out in the ECEPP/3 force field. Since the simple energy minimization performed for the early-stage structural forms of several proteins was found to be insufficient to approach the native-like structure [49–51], the introduction of an external hydrophobic force field

Table 1

General characteristics of early-stage (initial), late-stage (final), and native form of BPTI calculated as described in Section 2. The numbers of native R–R contacts reproduced in the early- as well as in the late-stage structural form are bracketed. The values of R_g and R_h given in parentheses are expected values calculated for different conformational states

Conformational state	RMSD (Å)	R–R	ASA _{tot} (Å ²)	ASA _p (Å ²)	ASA _H (Å ²)	R_g (Å)	R_h (Å)
Early-stage (in silico)	13.60	353 (265)	5066.55	1883.83	3182.72	16.31 (17.76)	18.06 (18.90)
Late-stage (in silico)	10.22	618 (344)	4460.14	1599.97	2860.17	12.12 (10.54)	15.83 (16.31)
Native		683	4018.73	1932.37	2086.36	12.41 (10.54)	16.11 (13.59)

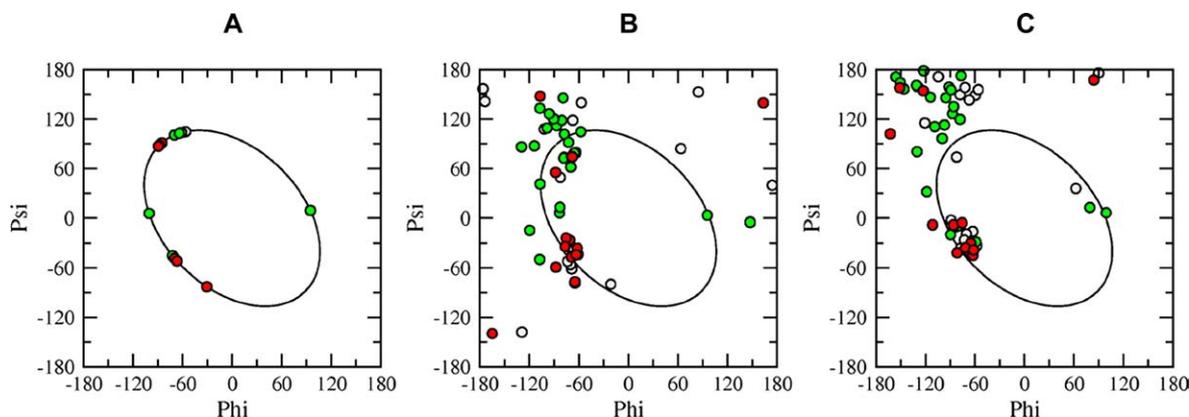


Fig. 3. Distribution of ϕ , ψ dihedral angles for early-stage (A), late-stage (B), and native (C) structural form of BPTI in Ramachandran plots. Residues of central β -sheet and C-terminal α -helix are colored green and red, respectively. Ellipse-path represents the limited conformational sub-space for early-stage folding (in silico).

in the form of *fuzzy-oil-drop* helped overcome the problem of the multidimensional energy surface.

3.5. Spatial distribution of $C\alpha$ atoms versus the geometrical center

The method applied for model assessment in CASP projects [80] was also used in this work to estimate the correctness of our models. The $D_{\text{center-C}\alpha}$ profiles for early-stage (initial), late-stage (final) and native structure are shown in Fig. 4. Structural similarity may be indicated by overlapping the lines representing the compared structures. The parallel orientation of profiles is interpreted as a similarity of structural forms in compared molecules oriented differently in the space. The increase of 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 the polypeptide chain in native and late-stage conformational states can be observed for residues 13–35 and 44–58. Those fragments correspond to the main structural elements: the central

antiparallel β -sheet and the C-terminal α -helix, respectively. However, the profiles revealed several key residues, particularly residues 7–12 and 36–43, responsible for different arrangements of polypeptide chains in both states (Fig. 2). It is noteworthy that the previous study on early-stage structural form of BPTI revealed that a significant structural change in exactly those fragments is necessary to attempt to explain the crystal structure [49]. Moreover, the regions where discrepancies between native and late-stage $D_{\text{center-C}\alpha}$ profiles are observed were found to be the most variable regions of BPTI [79,81,82].

3.6. Residue–residue interactions

Residue–residue interactions present in all discussed structural forms of BPTI are shown as contact maps in Fig. 5. 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 $C\alpha$ atoms were within 14 Å of one an-

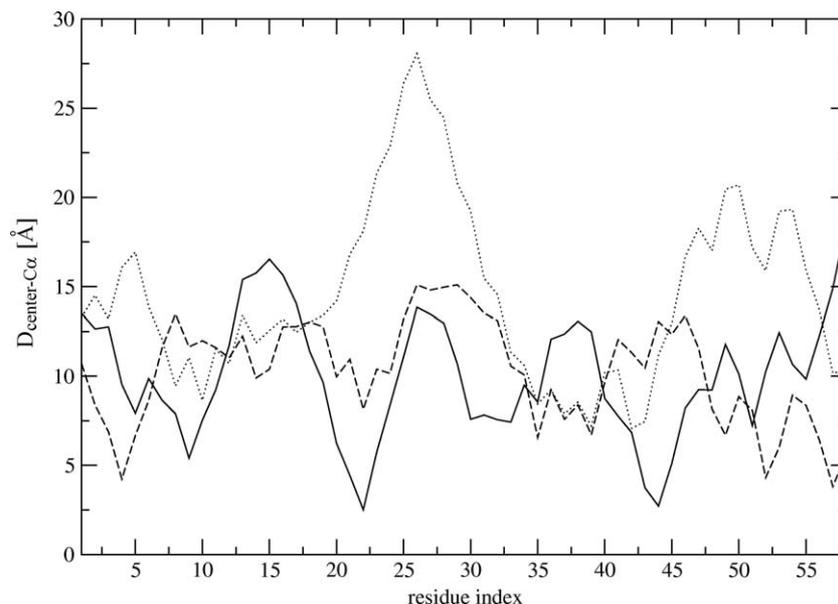


Fig. 4. Profiles of $D_{\text{center-C}\alpha}$ vectors for early-stage (dotted line), late-stage (dashed line) and native (solid line) structure of BPTI.

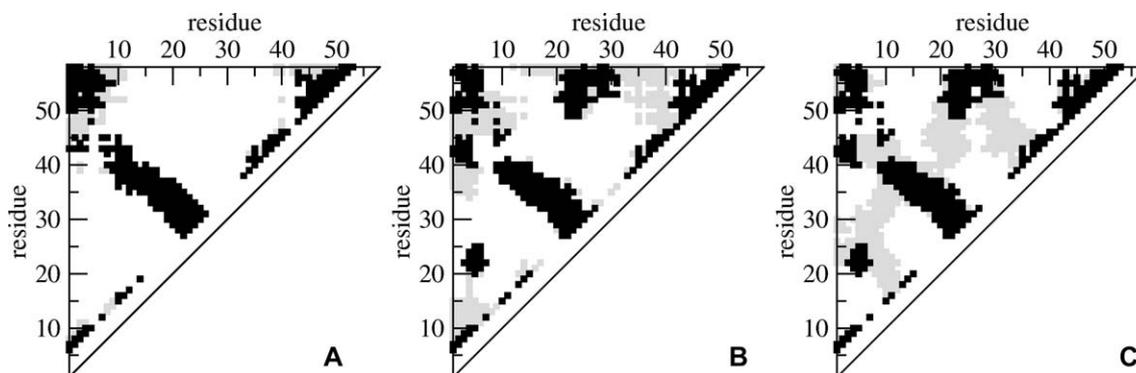


Fig. 5. Residue–residue (R–R) contact maps for BPTI: early-stage (A), late-stage (B), and native form (C). Gray squares express the complete set of R–R interactions found in a particular structural form, whereas black squares in A and B visualize the native R–R contacts reproduced in early- and late-stage form, respectively. Black squares presented in C depict the R–R contacts common for both late-stage and native structure.

other. BPTI is one of the smallest proteins in which the hydrophobic β -sheet has been claimed to play an important role in its folding [83,84]. The central β -sheet present already in the early-stage form is represented by the strong counter-diagonal of Fig. 5A, while the C-terminal α -helix appears as the upper right thick pattern. A similar contact pattern was observed at the early stages (1.3×10^{-3} s from initiation of folding) of protein folding dynamics based on a coarse-grained model [85] as well as for the results of molecular dynamics simulation of temperature-induced unfolding [79]. The final contact pattern (Fig. 5B) reproduces most of the important features of the native folded structure (Fig. 5C).

The total number of residue–residue interactions (R–R) for early-stage (initial), late-stage (final) and native structural form of BPTI was found to be 353, 618 and 683, respectively (Table 1). During the late-stage folding simulation the total number of residue–residue contacts significantly increased, but did not reach the amount characteristic of native structure. Furthermore, the percentage of the reproduced native R–R interactions grew 11.57%, to 344 (late-stage form) from 265 (early-stage form).

3.7. Solvent-exposed surface and radius of gyration analysis

The hydrophobic effect, causing non-polar side-chains to tend to cluster together in the protein interior, can be measured by the solvent-ASA of a protein. The radius of gyration (R_g) and hydrodynamic radius (R_h) estimate the characteristic volume of a globular protein and provides quantitative information on its compactness. Controlling the changes in the molecular dimensions is crucial for modeling the process of protein folding [86–89]. The properties of all discussed structural forms of BPTI displayed by ASA, R_g and R_h are presented in Table 1. Since the volume of *fuzzy-oil-drop* decreased, the packing density of simulated protein increased. The total and non-polar ASA of the unfolded early-stage form was found to be 26% and 53% greater than the native state, respectively. The radius of gyration calculated for the early-stage form was 45% greater than that calculated for the crystal structure. Similar characteristics of the unfolded state of BPTI were reported from molecular dynamics simulations of reduced BPTI in high

temperature [81] as well as pulsed field gradient NMR experiments [90]. Late-stage simulated folding significantly decreased the total, polar and non-polar accessible solvent area, reflecting the conformational rearrangement to achieve a stable, compact structure. Moreover, high accordance was found between the radii of gyration calculated for late-stage (final) and native structure of BPTI. A model presented in this paper was also validated by the analysis of how close modeled intermediates are approaching the hydrodynamic volumes experimentally determined at the different stages of folding. Table 1 shows that dimensions of both modeled intermediates are relatively close to the expected values calculated for different conformational states using a set equations from [55] and [87] for R_g and R_h , respectively.

3.8. Molecular dynamics simulation

The late-stage structural form of BPTI (as received according to folding procedure presented in this paper) was subjected to 1 ns molecular dynamics simulation in explicit solvent. The time series of potential and kinetic energies as well as RMSD values computed from the atomic trajectories for MD production run at 300 K are shown in Fig. 6. The simulation seems to be fairly stable, the potential and kinetic energies were fluctuating around a constant mean value (Fig. 6A). The structure has deviated from starting structure for 3.45 Å (Fig. 6B) likely due to an adaptation to the force field, indicate that a steady state characterized by stochastic oscillations was achieved. However, no conformational changes toward the native structural form were observed.

4. Conclusion

The early-stage folding (in silico) model was introduced to enable creation of the initial structural form for structure optimization procedure. The structure optimization in our model is driven by the hydrophobic interaction inside the *fuzzy-oil-drop* expressed by a three-dimensional Gaussian function. The three variables present in the three-dimensional Gaussian function represent the Cartesian distances. The parameter interpreted as mean value (all three mean values are common and equal

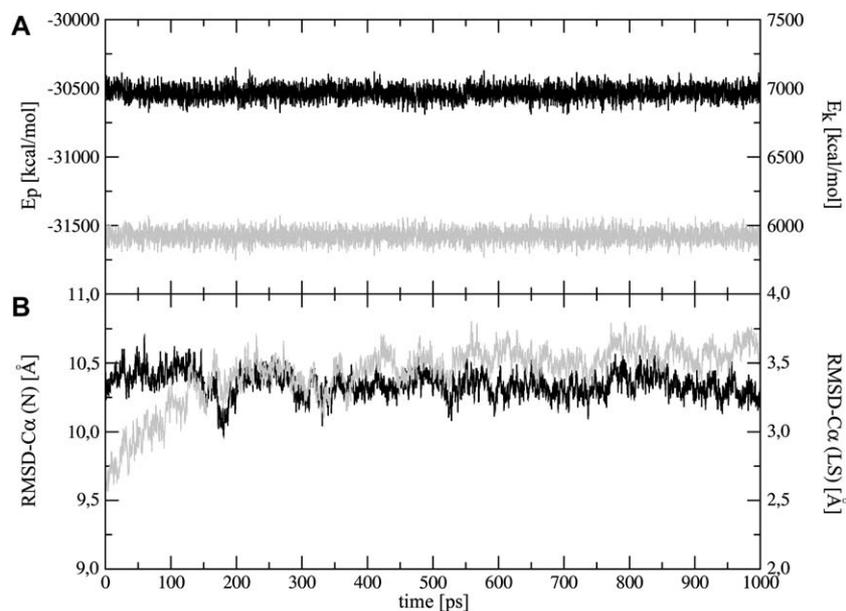


Fig. 6. The potential (E_p , black line) and kinetic (E_k , gray line) energy (A) as well as the RMSD-C α for crystal (N, black line) and late-stage (LS, gray line) structure (B) as a function of simulation time for the MD production run.

to 0.0) is localized at the origin of the coordinate system. This drop is called ‘fuzzy’ in our approach according to Gaussian function characteristics, where the probability—hydrophobicity density in our model—decreases in a distance-dependent manner. The *fuzzy-oil-drop* elastically changes its shape and size during the simulated late-stage folding (in silico). This change process is continuous in nature.

The early-stage folding (in silico) model, based on the backbone conformation together with sequence-to-structure information stored in a contingency table, and the *fuzzy-oil-drop* model assumed to represent hydrophobic collapse seem to work well all together, giving new insight into the nature of the protein folding process (in silico so far). Moreover our results seem to be consistent with available experimental observations and various folding/unfolding simulations.

The disulfide bonds were absent in the simulation. Their role can be critical for the folding process causing much better approach of the final structure to the native one (paper submitted for publication). The late-stage folding simulations of ribonuclease, lysozyme, hemoglobin and hypothetical membrane protein - target protein in CASP6 (TA0354_69_121) according to *fuzzy-oil-drop* model are presented elsewhere [55,91]. The early-stage model of TA0354_69_121 was submitted by our group in the CASP6 experiment. It should be noted that the ranking provided by CASP organizers revealed the fairly high accuracy of this prediction. The application of the *fuzzy-oil-drop* model as the continuation of *blind* prediction significantly improved the accuracy of structure prediction for this target protein [91].

The main advantage of the presented model is its universality, although some additional improvements are still necessary. So far the model presented in this paper was applied to single-domain globular proteins of up to 150 amino acids long. Large multi-domain proteins including several well-separated hydrophobic cores may be simulated using the bunch of coop-

erative *fuzzy-oil-drops*. The inverse function expressed as $1 - \Delta\hat{H}$ may be applied for “inside out” integral membrane proteins pushing the hydrophobic residues to be exposed toward the membrane. Moreover, any non-Gaussian function (e.g. Lorentz) can be also used for the theoretical *fuzzy-oil-drop* creation. The possibility of the *fuzzy-oil-drop* model application in such cases will be verified in the close future. The pattern of simulating the presence of an external force field seems to be generalized. Any process conditioned by the presence of an external force field of known (or assumed) form can be treated numerically as presented in this model.

The presented simulations (optimization and molecular dynamics simulation) were oriented on the verification of the role of hydrophobic core in folding process and structural stability. The high stability (despite of absence of SS-bonds) of received structural form in molecular dynamics simulation proves the model to be a tool to create the very well stabilizing hydrophobic core. According to expectations, the protein surface is covered by the hydrophilic residues which results as high solubility and stability in water environment.

The model is planned to be applied for blind prediction in CASP7 (summer 2006). It will give the good opportunity for model verification based on the larger spectrum of different proteins.

Acknowledgements

Many thanks to Professor Marek Pawlikowski (Faculty of Chemistry, Jagiellonian University) for 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.

Appendix A

A.1. Theoretical fuzzy-oil-drop

The *fuzzy-oil-drop* representing the environment for polypeptide folding is described by a three-dimensional Gauss function. The function usually interpreted as a probability distribution is assumed to represent the hydrophobicity distribution. If the j th point described by Cartesian coordinates (x_j, y_j, z_j) belongs to a box 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:

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

where $\sigma_x, \sigma_y, \sigma_z$ denote standard deviations and point $(\bar{x}, \bar{y}, \bar{z})$ represents the highest hydrophobicity value and occupies the position at the box center $(0, 0, 0)$. $\tilde{H}t_{sum}$ 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 1.0 for $(\bar{x}, \bar{y}, \bar{z})$, can be calculated.

A.2. Observed fuzzy-oil-drop

The observed hydrophobicity distribution within the *fuzzy-oil-drop* is calculated using the simple sigmoid function previously proposed to quantitatively describe the hydrophobic interactions [45]. The j th point collects hydrophobicity $\tilde{H}o_j$ as follows:

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

where N is the total number of residues in the protein under consideration, \tilde{H}_i^r denotes the hydrophobicity of the i th residue according to the normalized scale of hydrophobicity for amino acids, 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 9.0 Å following the original paper [45]. This means that only residues with $r_{ij} \leq c$ influence the j th point. $\tilde{H}o_{sum}$ is the sum of observed hydrophobicity for all analyzed grid points. Using $\frac{1}{\tilde{H}o_{sum}}$ as a normalizing coefficient the observed hydrophobicity can be compared to the theoretical hydrophobicity described previously.

A.3. The molecular dimensions of proteins in early-stage and native conformational state

The analysis of the size of rectangular box covering completely the protein molecule in its native and early-stage conformational state as dependent on number of amino acids in polypeptide chain revealed the degree of expected drop com-

pression during the implosion of hydrophobic collapse. Following correlations between the box volume covering the complete molecule (V) and chain length (N) were found on the basis of single-domain proteins analysis:

- $\log V = 3.5671 + 0.7725 \times \log N$ for native conformational state;
- $\log V = 3.0013 + 1.2271 \times \log N$ for early-stage conformational state.

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.

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.

References

- [1] R.L. Baldwin, Pieces of the folding puzzle, *Nature* 346 (1990) 409–410.
- [2] K.A. Dill, Theory for the folding and stability of globular proteins, *Biochemistry* 24 (1985) 1501–1509.
- [3] M. Levitt, A. Warshel, Computer simulation of protein folding, *Nature* 253 (1975) 694–698.
- [4] P.S. Kim, R.L. Baldwin, Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding, *Annu. Rev. Biochem.* 51 (1982) 459–489.
- [5] M. Karplus, D.L. Weaver, Diffusion–collision model for protein folding, *Biopolymers* 18 (1979) 1421–1437.
- [6] M. Karplus, D.L. Weaver, Protein-folding dynamics, *Nature* 260 (1976) 404–406.
- [7] V.R. Agashe, M.C. Shastry, J.B. Udgaonkar, Initial hydrophobic collapse in the folding of barstar, *Nature* 377 (1995) 754–757.
- [8] T. Sivaraman, T.K. Kumar, Y.T. Tu, W. Wang, W.Y. Lin, H.M. Chen, C. Yu, Secondary structure formation is the earliest structural event in the refolding of an all beta-sheet protein, *Biochem. Biophys. Res. Commun.* 260 (1999) 284–288.
- [9] Z. Guo, C.L. Brooks 3rd, E.M. Boczeko, Exploring the folding free energy surface of a three-helix bundle protein, *Proc. Natl. Acad. Sci. USA* 94 (1997) 10161–10166.
- [10] J.N. Onuchic, N.D. Socci, Z. Luthey-Schulten, P.G. Wolynes, Protein folding funnels: the nature of the transition state ensemble, *Fold. Des.* 1 (1996) 441–450.
- [11] H.S. Chan, K.A. Dill, Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics, *Proteins* 30 (1998) 2–33.
- [12] J.D. Bryngelson, P.G. Wolynes, Spin glasses and the statistical mechanics of protein folding, *Proc. Natl. Acad. Sci. USA* 84 (1987) 7524–7528.
- [13] W. Kauzmann, Some factors in the interpretation of protein denaturation, *Adv. Protein Chem.* 14 (1959) 1–63.
- [14] K.A. Dill, Dominant forces in protein folding, *Biochemistry* 29 (1990) 7133–7155.
- [15] R.R. Matheson, H.A. Scheraga, A method for predicting nucleation sites for protein folding based on hydrophobic contacts, *Macromolecules* 11 (1978) 819–829.
- [16] G.D. Rose, A.R. Geselowitz, G.J. Lesser, R.H. Lee, M.H. Zehfus, Hydrophobicity of amino acid residues in globular proteins, *Science* 229 (1985) 834–838.
- [17] R.L. Baldwin, Making a network of hydrophobic clusters, *Science* 295 (2002) 1657–1658.

- [18] C. Chothia, Structural invariants in protein folding, *Nature* 254 (1975) 304–308.
- [19] M.H. Klapper, On the nature of the protein interior, *Biochim. Biophys. Acta* 229 (1971) 557–566.
- [20] I.M. Klotz, Comparison of molecular structures of proteins: helix content; distribution of apolar residues, *Arch. Biochem. Biophys.* 138 (1970) 704–706.
- [21] J. Kyte, R.F. Doolittle, A simple method for displaying the hydrophobic character of a protein, *J. Mol. Biol.* 157 (1982) 105–132.
- [22] H. Meirovitch, H.A. Scheraga, 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 (1980) 1406–1414.
- [23] G.D. Rose, S. Roy, Hydrophobic basis of packing in globular proteins, *Proc. Natl. Acad. Sci. USA* 77 (1980) 4643–4647.
- [24] R. Bonneau, C.E. Strauss, D. Baker, Improving the performance of Rosetta using multiple sequence alignment information and global measures of hydrophobic core formation, *Proteins* 43 (2001) 1–11.
- [25] H. Holm, C. Sander, Evaluation of protein models by atomic solvation preference, *J. Mol. Biol.* 225 (1992) 93–105.
- [26] J. Novotny, A.A. Rashin, R.E. Bruccoleri, Criteria that discriminate between native proteins and incorrectly folded models, *Proteins* 4 (1988) 19–30.
- [27] R.B. Zhou, B.D. Silverman, A.K. Royyuru, P. Athma, Spatial profiling of protein hydrophobicity: native vs. decoy structures, *Proteins* 52 (2003) 561–572.
- [28] J. Lee, S.M. Shin, Understanding β -hairpin formation by molecular dynamics simulation of unfolding, *Biophys. J.* 81 (2001) 2507–2516.
- [29] C.J. Tsai, R. Nussinov, Hydrophobic folding units derived from dissimilar monomer structures and their interactions, *Protein Sci.* 6 (1997) 24–42.
- [30] B. Zagrovic, E.J. Sorin, V.S. Pande, β -Hairpin folding simulations in atomistic details using an implicit solvent model, *J. Mol. Biol.* 313 (2001) 151–169.
- [31] B.D. Silverman, Hydrophobic moments of protein structures: spatially profiling the distribution, *Proc. Natl. Acad. Sci. USA* 98 (2001) 4996–5001.
- [32] B.D. Silverman, A two-component nucleation model of protein hydrophobicity, *J. Theor. Biol.* 216 (2002) 139–146.
- [33] B.D. Silverman, Hydrophobicity of transmembrane proteins: spatially profiling the distribution, *Protein Sci.* 12 (2003) 586–599.
- [34] V.I. Abkevich, A.M. Gutin, E.I. Shakhnovich, Specific nucleus as the transition state for protein folding: evidence from the lattice model, *Biochemistry* 33 (1994) 10026–10036.
- [35] V.S. Pande, D.S. Rokhsar, Folding pathway of a lattice model for proteins, *Proc. Natl. Acad. Sci. USA* 96 (1999) 1273–1278.
- [36] A. Sali, E. Shakhnovich, M. Karplus, Kinetics of protein folding. A lattice model study of the requirements for folding to the native state, *J. Mol. Biol.* 235 (1994) 1614–1636.
- [37] J. Skolnick, A. Kolinski, Dynamic Monte Carlo simulations of a new lattice model of globular protein folding, structure and dynamics, *J. Mol. Biol.* 221 (1991) 499–531.
- [38] H. Taketomi, Y. Ueda, N. Go, 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 (1975) 445–459.
- [39] K.A. Dill, S. Bromberg, K. Yue, K.M. Fiebig, D.P. Yee, P.D. Thomas, H.S. Chan, Principles of protein folding—a perspective from simple exact models, *Protein Sci.* 4 (1995) 561–602.
- [40] C.M. Dobson, The structural basis of protein folding and its links with human disease, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356 (2001) 133–145.
- [41] A. Kolinski, D. Gront, P. Pokarowski, J. Skolnick, A simple lattice model that exhibits a protein-like cooperative all-or-none folding transition, *Biopolymers* 69 (2003) 399–405.
- [42] A. Kolinski, J. Skolnick, Monte Carlo simulations of protein folding. I. Lattice model and interaction scheme, *Proteins* 18 (1994) 338–352.
- [43] D. Baker, A surprising simplicity to protein folding, *Nature* 405 (2000) 39–42.
- [44] P. Pokarowski, A. Kolinski, J. Skolnick, A minimal physically realistic protein-like lattice model: designing an energy landscape that ensures all-or-none folding to a unique native state, *Biophys. J.* 84 (2003) 1518–1526.
- [45] M. Levitt, A simplified representation of protein conformations for rapid simulation of protein folding, *J. Mol. Biol.* 104 (1976) 59–107.
- [46] W. Jurkowski, M. Brylinski, L. Konieczny, Z. Wiiniowski, I. Roterman, Conformational subspace in simulation of early-stage protein folding, *Proteins* 55 (2004) 115–127.
- [47] I. Roterman, Modelling the optimal simulation path in the peptide chain folding—studies based on geometry of alanine heptapeptide, *J. Theor. Biol.* 177 (1995) 283–288.
- [48] I. Roterman, The geometrical analysis of peptide backbone structure and its local deformations, *Biochimie* 77 (1995) 204–216.
- [49] M. Brylinski, W. Jurkowski, L. Konieczny, I. Roterman, Limited conformational space for early-stage protein folding simulation, *Bioinformatics* 20 (2004) 199–205.
- [50] W. Jurkowski, M. Brylinski, L. Konieczny, I. Roterman, Lysozyme folded in silico according to the limited conformational sub-space, *J. Biomol. Struct. Dyn.* 22 (2004) 149–158.
- [51] M. Brylinski, W. Jurkowski, L. Konieczny, I. Roterman, Limitation of conformational space for proteins—early stage folding simulation of human α and β hemoglobin chains, *TASK Quarterly* 8 (2004) 413–422.
- [52] S. Hayward, Peptide-plane flipping in proteins, *Protein Sci.* 10 (2001) 2219–2227.
- [53] M. Brylinski, L. Konieczny, P. Czerwonko, W. Jurkowski, I. Roterman, Early-stage folding in proteins (in silico) sequence-to-structure relation, *J. Biomed. Biotechnol.* 2 (2005) 65–79.
- [54] M. Brylinski, L. Konieczny, I. Roterman, SPI—structure predictability index for protein sequences, *In Silico Biol.* 5 (2004) 0022.
- [55] M. Brylinski, L. Konieczny, I. Roterman, Fuzzy-oil-drop hydrophobic force field—a model to represent late-stage folding (in silico) of lysozyme, *J. Biomol. Struct. Dyn.* 23 (2006) 519–528.
- [56] H.H. Rosenbrock, An automatic method for finding the greatest or least value of a function, *Comput. J.* 3 (1960) 175–184.
- [57] M.J. Sippl, G. Nemethy, H.A. Scheraga, Intermolecular potentials from crystal data. 6. Determination of empirical potentials for O—H:O=C hydrogen bonds from packing configurations, *J. Phys. Chem.* 88 (1984) 6231–6233.
- [58] F.A. Momany, R.F. McGuire, A.W. Burgess, H.A. Scheraga, Energy parameters in polypeptides. VII. Geometric parameters, partial charges, non-bonded interactions, hydrogen bond interactions and intrinsic torsional potentials for naturally occurring amino acids, *J. Phys. Chem.* 79 (1975) 2361–2381.
- [59] L.G. Dunfield, A.W. Burgess, H.A. Scheraga, Energy parameters in polypeptides. 8. Empirical potential energy algorithm for the conformational analysis of large molecules, *J. Phys. Chem.* 82 (1978) 2609–2616.
- [60] G. Nemethy, M.S. Pottle, H.A. Scheraga, Energy parameters in polypeptides. 9. Updating of geometrical parameters, non-bonding interactions and hydrogen bonding interactions for naturally occurring amino acids, *J. Phys. Chem.* 87 (1983) 1883–1887.
- [61] G. Nemethy, K.D. Gibson, K.A. Palmer, C.N. Yoon, G. Paterlini, A. Zagari, S. Rumsey, H.A. Scheraga, 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 (1992) 6472–6484.
- [62] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.R. Ross, T.E. Cheatham, I.S. DeBolt, D. Ferguson, G. Seibel, P.A. Kollman, 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 (1995) 1–41.
- [63] J. Wang, P. Cieplak, P.A. Kollman, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?, *J. Comput. Chem.* 21 (2000) 1049–1074.
- [64] J.P. Ryckaert, G. Cicotti, H.J.C. Berendsen, Numerical integration of the Cartesian equation of motion of a system constraints: molecular dynamics of N-alkanes, *J. Comput. Phys.* 23 (1977) 327–341.

- [65] O.V. Tsodikov, M.T. Record Jr., Y.V. Sergeev, Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature, *J. Comput. Chem.* 23 (2002) 600–609.
- [66] J. Garcia De La Torre, M.L. Huertas, B. Carrasco, Calculation of hydrodynamic properties of globular proteins from their atomic-level structure, *Biophys. J.* 78 (2000) 719–730.
- [67] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graph* 14 (1996) 33–38 (27–38).
- [68] C. Bystroff, Y. Shao, Fully automated ab initio protein structure prediction using I-SITES, HMMSTR and ROSETTA, *Bioinformatics* 18 (Suppl. 1) (2002) S54–S61.
- [69] A.G. de Brevern, C. Benros, R. Gautier, H. Valadie, S. Hazout, C. Etchebest, Local backbone structure prediction of proteins, *In Silico Biol.* 4 (2004) (0031).
- [70] A.G. de Brevern, S. Hazout, ‘Hybrid protein model’ for optimally defining 3D protein structure fragments, *Bioinformatics* 19 (2003) 345–353.
- [71] J.S. Fetrow, M.J. Palumbo, G. Berg, Patterns, structures, and amino acid frequencies in structural building blocks, a protein secondary structure classification scheme, *Proteins* 27 (1997) 249–271.
- [72] K.F. Han, D. Baker, Global properties of the mapping between local amino acid sequence and local structure in proteins, *Proc. Natl. Acad. Sci. USA* 93 (1996) 5814–5818.
- [73] S. Gnanakaran, A.E. Garcia, Helix-coil transition of alanine peptides in water: force field dependence on the folded and unfolded structures, *Proteins* 59 (2005) 773–782.
- [74] D. Paschek, S. Gnanakaran, A.E. Garcia, Simulations of the pressure and temperature unfolding of an alpha-helical peptide, *Proc. Natl. Acad. Sci. USA* 102 (2005) 6765–6770.
- [75] B. Rost, C. Sander, Prediction of protein secondary structure at better than 70% accuracy, *J. Mol. Biol.* 232 (1993) 584–599.
- [76] A. Zemla, C. Venclovas, K. Fidelis, B. Rost, A modified definition of Sov, a segment-based measure for protein secondary structure prediction assessment, *Proteins* 34 (1999) 220–223.
- [77] T. Nakazawa, H. Kawai, Y. Okamoto, M. Fukugita, β -Sheet folding of fragment (16–36) of bovine pancreatic trypsin inhibitor as predicted by Monte Carlo simulated annealing, *Protein Eng.* 5 (1992) 495–503.
- [78] K.C. Chou, G. Nemethy, M.S. Pottle, H.A. Scheraga, Folding of the twisted beta-sheet in bovine pancreatic trypsin inhibitor, *Biochemistry* 24 (1985) 7948–7953.
- [79] V. Daggett, M. Levitt, Protein unfolding pathways explored through molecular dynamics simulations, *J. Mol. Biol.* 232 (1993) 600–619.
- [80] C.A. Orengo, J.E. Bray, T. Hubbard, L. LoConte, I. Sillitoe, Analysis and assessment of ab initio three-dimensional prediction, secondary structure, and contacts prediction, *Proteins Suppl* 3 (1999) 149–170.
- [81] S.L. Kazmirski, V. Daggett, Simulations of the structural and dynamical properties of denatured proteins: the “molten coil” state of bovine pancreatic trypsin inhibitor, *J. Mol. Biol.* 277 (1998) 487–506.
- [82] A.A. Kossiakoff, M. Randal, J. Guenot, C. Eigenbrot, Variability of conformations at crystal contacts in BPTI represent true low energy structure-correspondence among lattice packing and molecular dynamics structures, *Proteins* 14 (1992) 65–74.
- [83] M. Dadlez, Hydrophobic interactions accelerate early stages of the folding of BPTI, *Biochemistry* 36 (1997) 2788–2797.
- [84] T.G. Oas, P.S. Kim, A peptide model of a protein folding intermediate, *Nature* 336 (1988) 42–48.
- [85] A. Fernandez, K. Kostov, R.S. Berry, From residue matching patterns to protein folding topographies: general model and bovine pancreatic trypsin inhibitor, *Proc. Natl. Acad. Sci. USA* 96 (1999) 12991–12996.
- [86] C. Tanford, Protein denaturation, *Adv. Protein Chem.* 23 (1968) 121–282.
- [87] O. Tcherkasskaya, E.A. Davidson, V.N. Uversky, Biophysical constraints for protein structure prediction, *J. Proteome Res.* 2 (2003) 37–42.
- [88] O. Tcherkasskaya, V.N. Uversky, Denatured collapsed states in protein folding: example of apomyoglobin, *Proteins* 44 (2001) 244–254.
- [89] D.K. Wilkins, S.B. Grimshaw, V. Receveur, C.M. Dobson, J.A. Jones, L.J. Smith, Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques, *Biochemistry* 38 (1999) 16424–16431.
- [90] H. Pan, G. Barany, C. Woodward, Reduced BPTI is collapsed. A pulse field gradient NMR study of unfolded and partially folded bovine pancreatic trypsin inhibitor, *Protein Sci.* 6 (1997) 1985–1992.
- [91] L. Konieczny, M. Brylinski, I. Roterman, Gauss-function-based model of hydrophobicity density in proteins, *In Silico Biol.* 6 (2006) 0002.