Lysozyme Folded *In Silico* According to the Limited Conformational Sub-space

http://www.jbsdonline.com

W. Jurkowski[1,3]

M. Bryliński[1,3]

L. Konieczny[2]

I. Roterman[3]*

[1]Faculty of Chemistry

Jagiellonian University

Ingardena 3

30-060 Krakow, Poland

[2]Institute of Medical Biochemistry

Collegium Medicum –

Jagiellonian University

Kopernika 7

31-034 Krakow, Poland

[3]Dept. of Bioinformatics and Telemedicine

Collegium Medicum –

Jagiellonian University

Kopernika 17

31-501 Krakow, Poland

Abstract

The conformational sub-space oriented on early-stage protein folding is applied to lysozyme folding. The part of the Ramachandran map distinguished on the basis of a geometrical model of the polypeptide chain limited to the mutual orientation of the peptide bond planes is shown to deliver the initial structure of the polypeptide for the energy minimization procedure in the *ab initio* model of protein folding prediction. Two forms of energy minimization and molecular dynamics simulation procedures were applied to the assumed early-stage protein folding of lysozyme. One of them included the disulphide bond system and the other excluded it. The post-energy-minimization and post-dynamics structures were compared using RMS-D and non-bonding contact maps to estimate the degree of approach to the native, target structure of the protein molecule obtained using the limited conformational sub-space for the early stage of folding.

Introduction

Two ideas are the background for the model presented in this paper.

One of them was put forward by Alonso and Daggett (1): "The importance of the pathways from unfolded to native states stems from the combinatorics of the protein folding problem; many small globular proteins unfold and refold reversibly to the same native structure, but they cannot search all possible conformations for the most stable state in a reasonable period of time. In light of these findings, there must be mechanism to limit the conformational search in particular regions of conformational space or to narrow the pathways between the unfolded and native states".

The second was expressed by Dobson (2): "The solution to this apparent paradox [Levinthal paradox] has emerged recently through the consideration of the energy surfaces or 'landscapes' on which the folding reaction occurs. This approach recognizes that folding should not be considered as breaking and forming a small number of single, strong covalent bonds, but as a biased search on an energy surface that is generally rather flat, *i.e.* where there are many conformations of similar energy. Such a surface arises because the conformation of a protein is determined by a very large number of relatively weak non-covalent interactions. Such model need to be sophisticated enough to encompass key features of protein folding *e.g.* there is a Levinthal paradox, but simple enough to allow many simulations to be carried out in a reasonable time".

Some simplified models have been proposed to solve this problem: the simplified structure of amino acid representation (3, 4) and limitation of the conformational space to four low-energy basins (5-7).

Phone/Fax: (0048) 124214057
Email: myroterm@cyf-kr.edu.pl
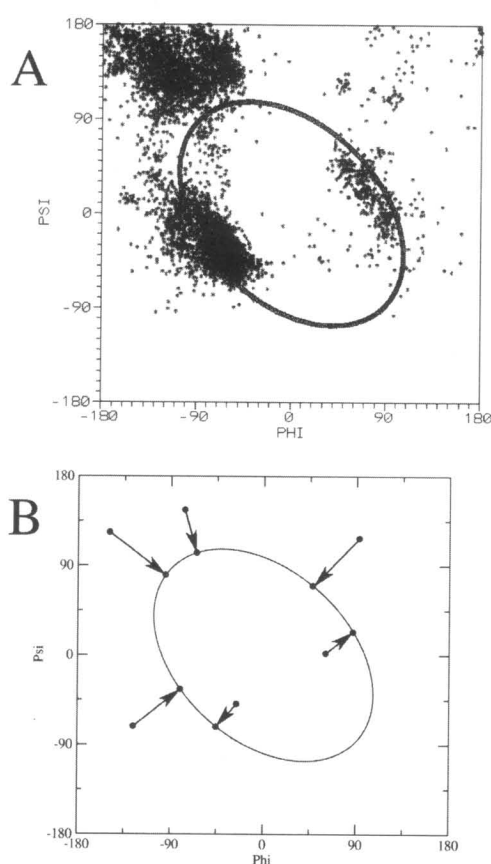
A simple model limiting the size of the stochastic search is proposed in this paper. The detailed background of the model was presented previously (8-10). The ellipse-limited path on the Ramachandran map is assumed to deliver the early-stage of folding. The ellipse path was distinguished based only on the mutual peptide bond plane orientation, which was found to be sufficient for structure classification (11). Side chain-side chain interactions were not present in the original model. This assumption is in accordance with suggestions that the backbone structure dominates in early-stage polypeptide structure formation and that side chain – side chain interaction occurs in later stages of polypeptide folding (12). Step-by-step introduction of this kind of interaction, which causes deviation from the original ellipse-path-limited conformational space, makes the successive structures of the polypeptide chain more and more similar to the target, native structure of the protein.

The lysozyme molecule was selected as a model because of its appropriate size (particularly for *ab initio* methods) and the mixed character of the ordered structural forms present in it. Moreover, the presence of disulphide bonds allows investigation of the influence of SS-bonds on the folding process.

Materials and Methods

Ellipse-path-derived Structure Creation

The lysozyme (2EQL according to PDB identification) structure was taken for analysis. The Phi, Psi angles as they appear in the native form of the proteins were calculated for each amino acid. Their values (Fig. 1A) were changed according to the criterion of shortest distance versus the ellipse-path Phi_{ell}, Psi_{ell} angles (Fig. 1B). The model leading to the ellipse-shaped conformational sub-space is presented in detail elsewhere (8-10). The ellipse-derived structures for lysozyme were formed according to the ECEPP/3 program. The Omega dihedral angles were taken as 180° for all amino acids. The side chain structures were created according to ECEPP/3 standards.

Energy Minimization Procedure

The energy minimization procedure was performed using the ECEPP/3 program for the ellipse-derived structure of lysozyme (13). The Phi, Psi angles were calculated for post-minimization structures. The unconstrained minimization solver with analytical gradient was used. The values of absolute and relative function convergence tolerances were set at 1×10^{-3} and 1×10^{-5}, respectively. The energy minimization procedure was carried out both with and without properly defined disulfide bonds. The coordinates and values of the backbone dihedral angles were saved for analysis at 10-step intervals. The energy minimization procedure for lysozyme was done on an SGI Origin 2000 in the computing center of TASK in Gdansk.

Molecular Dynamics Simulation Protocol

The molecular dynamics simulation starting with the ellipse-derived forms was performed using the AMBER 7.0 program (14) with ff99 force field (15). The water environment including the charge screening effects of salt was represented implicitly by the analytic Generalized Born (GB) model (16). Long range nonbonded interactions were truncated by using a 12 Å cutoff (electrostatic and VdW). Bond constraints were imposed on all bonds involving hydrogen atoms via the SHAKE algorithm (17) during simulation. The surrounding salt concentration was controlled at physiological conditions (0.2mol/l). The simulation procedures were conducted as follows:

I.   1000 steps of energy minimization. After 100 steps of steepest descent minimization the method was switched to the conjugant gradient method. All atoms in each residue were constrained to

Figure 1: (A) The ellipse-path limited conformational sub-space in relation to Phi, Psi angles as they appear in real proteins. (B) The shortest distance between particular Phi, Psi angles and the ellipse-belonging point represents the way in which the Phi, Psi angles on the ellipse can be found.

their original positions with a force constant k = 5.0 [kcal/mol/Å^2].

II.   10ps of equilibration with 2fs integration time-step, during which the temperature of the system was gradually raised from 0 to 300K. The molecules were constrained to their energy-minimized structure by a weaker potential of 1.0 [kcal/mol/Å^2]. At every 1000 steps the translational and rotational motion was removed.

III.   100 ps of unconstrained MD simulation at 300K with a 2 fs integration time-step.

IV.   4500 steps of unconstrained conjugant gradient minimization, followed by 500 steps of steepest descent minimization.

Steps III and IV were performed three times. The coordinates were saved at 100 step intervals for the analysis. All calculations using AMBER were done in the Department of Bioinformatics and Telemedicine of Collegium Medicum, Jagiellonian University.

Structure Comparison

The structures (native, elliptical, post-energy minimization and post-dynamics in two forms with and without SS-bonds) were compared using the following criteria:

I.   The distances between the geometric center of the molecule and sequential C_α atoms in the polypeptide chain were calculated. The profile of these distances for each amino acid revealed a rough degree of similarity. The polypeptide fragments distinguished according to the different forms of similarity between the profile of the native form and the others under consideration were also characterized by RMS-D calculation. The RMS-D values for selected fragments were calculated after individually overlapping the fragments taken from the native, ellipse-derived structural form (and other structural forms taken into account). The average value of RMS-D was calculated for each distinguished fragment.

II.   The number of non-bonding interactions was calculated for all structural forms, assuming a cutoff distance equal to 12 Å, and was compared to the number of these interactions in the native form of the protein.

III. The box large enough to cover the whole molecule was also calculated for each structural form of the protein molecule to express the size of the molecule.

The longest Cα-Cα distance was taken as the D_Z measure (distance along the Z-axis), the longest Cα-Cα distance in the XY-plane was taken as the D_Y measure (distance along the Y-axis), and the difference between the highest and lowest values of X was taken as the measure of the D_X box edge. The box volume (V) was expressed as $V = D_x \cdot D_y \cdot D_z$.

Results

Phi, Psi Dihedral Angle Changes

The Phi, Psi angle distribution exposing the range of dihedral angle change is shown for all discussed structural forms in Figure 2A for native form of lysozyme, Figure 2B for ellipse-derived structure, Figure 2C for post-energy-minimization structure without SS-bonds present, Figure 2D for post-energy-minimization structure with SS-bonds present, Figure 2E for the post-dynamics structure of lysozyme obtained taking the structure without disulphide bonds declared, and finally Figure 2F for the post-dynamics structure with SS-bonds declared. The similarity of the Phi, Psi angle distribution can be seen between the native structure and post-ener-

Figure 2: The structure of lysozyme and its Phi, Psi angle distribution versus the ellipse-path. (**A**) Native form, (**B**) ellipse-based structure, (**C**) post-energy-minimization structural form of lysozyme with SS-bonds absent in the procedure, (**D**) post-energy-minimization structural form of lysozyme with SS-bonds present in the procedure, (**E**) post-dynamics structural form of lysozyme with SS-bonds absent in the procedure, (**F**) post-dynamics structural form of post-energy-minimization structure of lysozyme with SS-bonds present in the procedure.
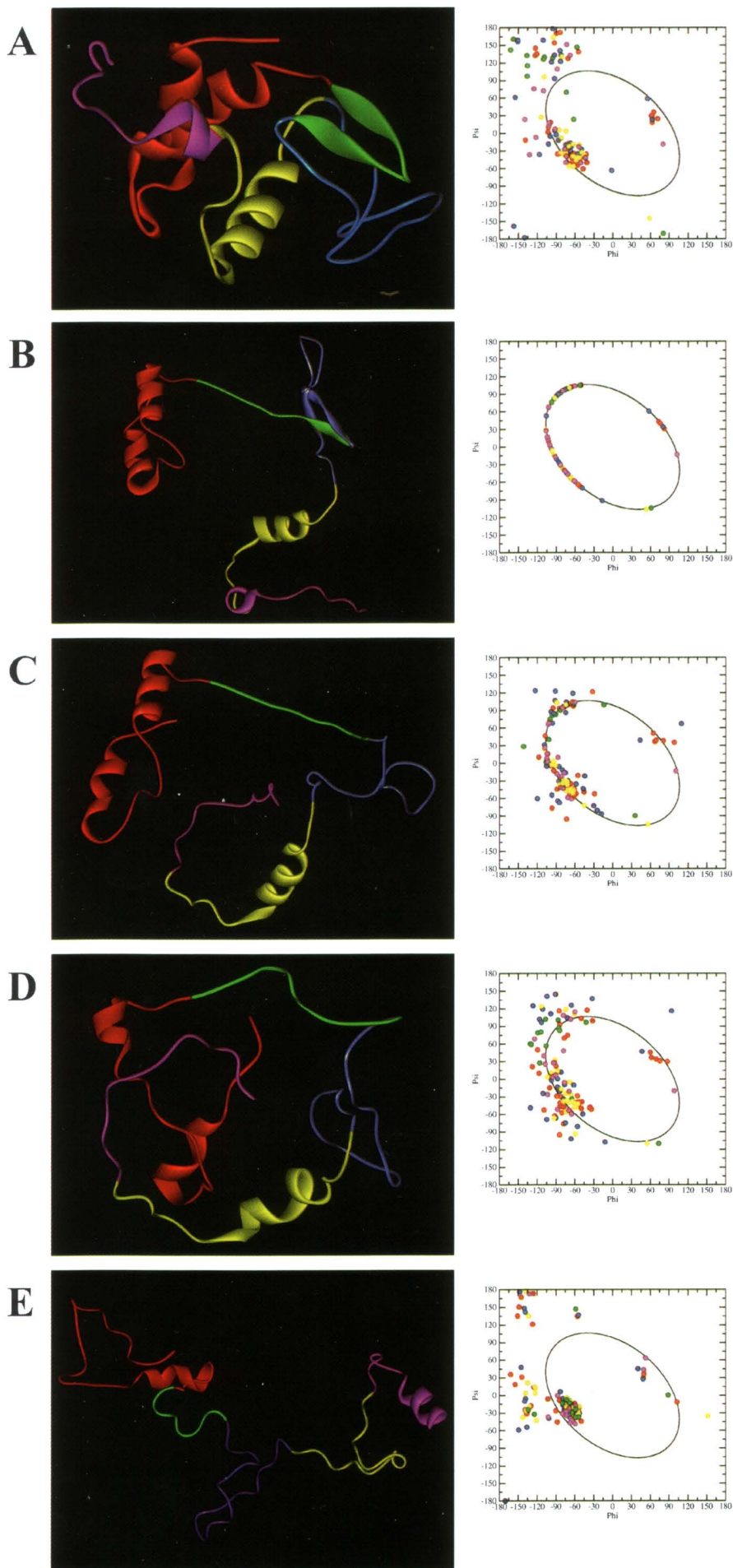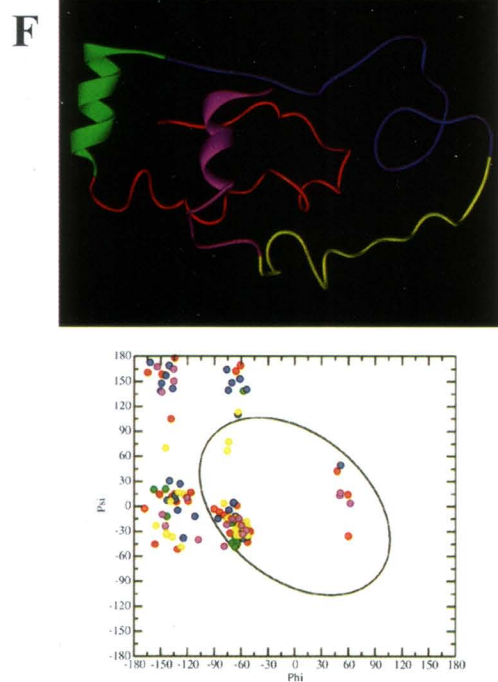
gy minimization structure with the SS-bond system introduced. The structure obtained after energy minimization with disulphide bonds declared was used in molecular dynamics simulation to reveal whether this simulation can cause it to approach the native structure of the protein.

Visual Analysis

Figure 2 visualizes structural changes in lysozyme in different conditions. Color notation differentiates particular polypeptide fragments and distinguishes them according to their similarity as measured by the $D_{center-C\alpha}$ vector profiles. The same color notation was used for the Phi, Psi angle distributions and RMS-D fragments.

No signs of approach toward native form can be seen in the energy minimization procedure performed without SS-bonds. A quite good approach was reached for the same procedure with SS-bonds defined according to the natural system present in this molecule.

Quite unusual and unexpected α-helical form shows up in post-dynamics structure with SS-bonds declared. This is the β-structural fragment in native structural form of the protein. The molecular dynamics simulation does not cause approach toward the native structure. The best structural form of lysozyme seems to be represented by the post-energy minimization form with the SS-bonds system declared.

Spatial Distribution of C_α Atoms Versus the Geometrical Center

The profile of the length of vectors linking the geometrical center with sequential C_α atoms gives insight into the three-dimensional relative displacements versus the native form of the protein (18). Structural similarity may be indicated by overlapping the lines representing the two compared structures. Parallel orientation of profiles is interpreted as similarity of structural forms in compared molecules oriented differently in the space. The increase of vector length versus native structure is obvious, due to the extension of the structure that is always associated with the transformation from native to ellipse-path-delimited structure. The profiles for all structural forms of lysozyme discussed in this paper are presented in Figure 3. Five fragments can be distinguished. The N-terminal fragment containing 1-42 aa is characterized by very similar regularities of minima and maxima dispersions in the $D_{center-C\alpha}$ profile in the compared structures (excluding ellipse-derived) and simultaneously by the lowest value of RMS-D. The highest similarity of profiles can be seen in fragments 56-85 aa and 86-110 aa, although fragment 56-85 aa represents higher RMS-D. The highest discrepancy in profiles is present in the 43-55 aa fragment.
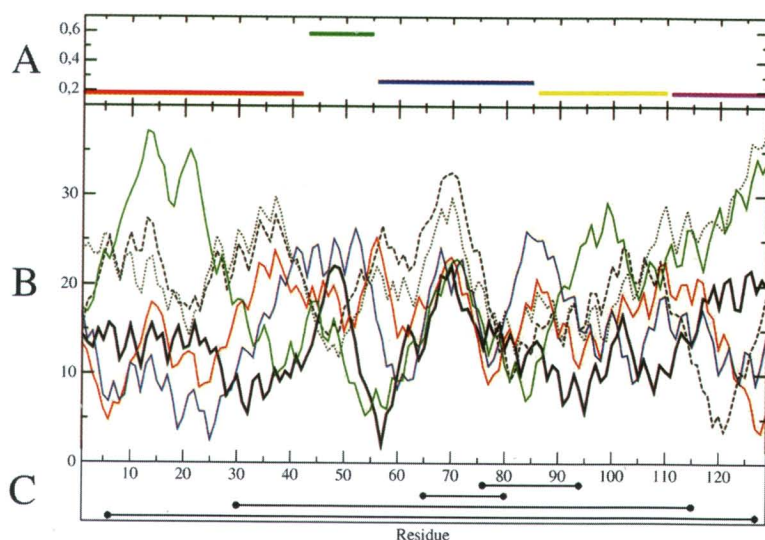

Figure 3: Comparison of structural forms of lysozyme. (**A**) RMS-D (per residue) calculated for structurally differentiated polypeptide fragments. The fragments were defined according to the profile presented in **B**. Parallel fragments of curves represent the correct spatial orientation of the polypeptide, while dissimilar regularity of curves represents fragments with low similarity of the spatial orientation of the particular polypeptide fragment. (**B**) Profile representing the distribution of distances linking the geometrical center of the molecule with sequential C_α atoms.

Continuous bold line, native form; Dotted line, ellipse-derived structure; Dashed line, post-energy minimization structure with SS-bonds absent; Continuous red line, post-energy minimization structure with SS-bonds present; Continuous green line, post-dynamics structure with SS-bonds absent; Continuous blue line, post-dynamics structure of post-energy minimization form with SS-bonds present.

(**C**) SS-bond system in lysozyme. The color notation introduced in all figures in this paper accords with the fragments distinguished in this figure.

The native non-bonding interactions present in all discussed structural forms are shown in Figure 4. Significant similarity of non-bonding contact distributions can be seen between native structure and the form with SS-bonds present during the energy minimization and molecular dynamics simulation procedures. The appearance of non-bonding contacts between the N-terminal and C-terminal fragments can be seen in Figure 4C and Figure 4F. The dispersion of non-bonding contacts in the post-dynamics structure (Fig. 4F) is even better in subjective evaluation.

Figure 4: Non-bonding contacts in lysozyme. (A) Native form, (B) ellipse-derived structure, (C) post-energy-minimization with SS-bonds absent, (D) post-energy minimization with SS-bonds present, (E) post-dynamics with SS-bonds absent, (F) post-dynamics of post-energy minimization form with SS-bonds present.
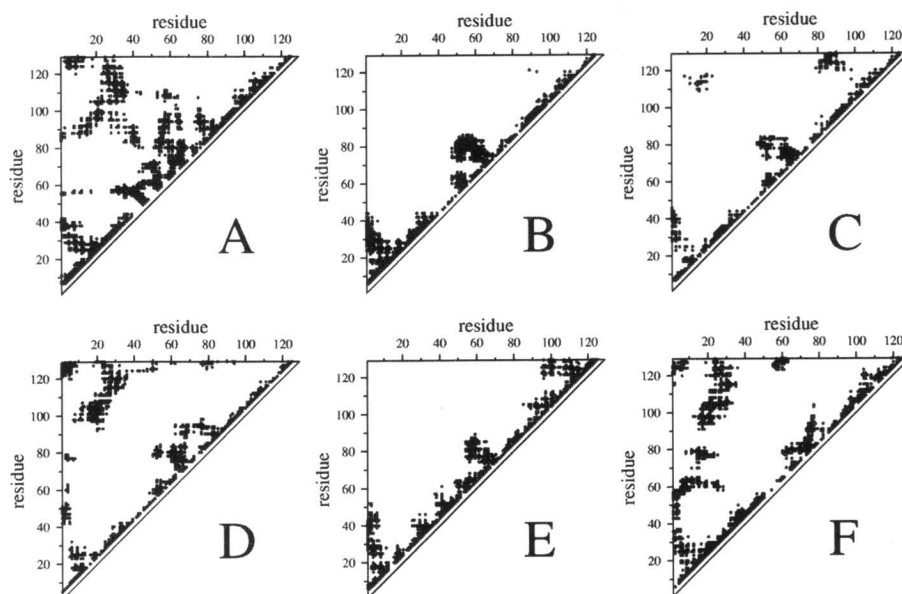
The percentages of native non-bonding interactions present in the ellipse-derived structure are as follows: 48.02% for ellipse-derived, 44.28% and 40.76 % for post-energy-minimization with and without SS-bonds, 44.41% and 43.77% for post-dynamics structures with and without SS-bonds, respectively. The highest percentage of native non-bonding contacts in ellipse-derived structure seems to be caused by the helical fragment present in the N-terminal fragment. This helical fragment seems to change its structure to a random one in post-energy-minimization procedures. This quantity obviously depends on the molecule under study (19) and is strongly related to the percentage of helical forms in the analyzed structure. The ellipse path passes through the region attributed to helical forms on the Ramachandran map, additionally explaining the rather high percentage of native-like non-bonding interactions.

Change of Molecular Size

The most critical problem concerning the relation between ellipse-derived structures and native structures is the question of how much the size of the molecule is changed, that is, what degree of compactness of ellipse-derived structures is necessary to reach the native form. The relative (versus the native form of the protein) increase of the box size (volume) containing the whole protein molecule (the size of which is calculated according to the procedure given in ***Methods***) appeared to be 2.39 for ellipse-derived structure, for post-energy-minimization with (1.95) and without SS-bonds (2.76) and post-dynamics (1.71 and 1.97) (with and without SS bonds declared respectively). The significant influence of the presence of SS-bonds on the box size is obvious. The post-dynamics structure represents the highest packing, expressed by only a 1.71 increase of size versus the native form.

Discussion

Experiments provide clear evidence for the existence of partially folded intermediates (28-31). Two intermediate structural forms are proposed for lysozyme folding

in Dobson (2). The first one leads to α-domain containing mostly helical fragments packed together. The second intermediate is represented by β-domain. The second intermediate differs only by its lower packing, while the structural shape of the molecule is very well defined. The results obtained in this work differ in the definition of the unfolded state (2) which represents only the random coiled form, while according to the model presented in this work helical fragments are present even in the early stage of folding.

Neither energy minimization nor molecular dynamics led to satisfactory similarity of the structure to the native one. One should mention, though, that the starting structure was applied to the simplest form of the procedures, performed in a standard way. The mostly β-structural fragments are far from expectations. This structural form causes problems, as was evident in the CASP competition (18). The appearance of helical fragment in post-dynamics form can be explained by the fact, that no β-form was produced in this area and in consequence the lowest energy structural form was produced.

The general interpretation of the results is that the molecular dynamics simulation did not correct the structure of the protein. Subjective estimation leads to the conclusion, that the best approach is represented by the post-energy minimization procedure with the SS-bond system present. Further attempts are in progress to estimate the influence of temperature, solvent model and longer molecular dynamics simulation. Visual analysis of the best structural form (post-energy minimization with the SS-bond system declared) is still far from satisfactory. The proper level of packing apparently was not reached. This is why introducing dynamics made the system even worse. The intermediate step representing hydrophobic collapse seems to be necessary, a model of which is under consideration and will be presented soon.

The aim of this paper was to present a model for formation of early-stage protein structure. Other proteins analyzed using the same model are: BPTI (20) and ribonuclease (10).

The stages distinguished as partitioning the protein folding process proposed by Fergusson and Fersht (21) are as follows: i, specific or nonspecific chain collapse; ii, formation of secondary and tertiary structure according to the balance of local and non-local, native and non-native interactions; and iii, desolvation of the chain as it folds to a lower energy conformation. Although the sequence of events may be dependent on the protein under study (19, 22), the chosen experimental conditions and perhaps the time regime of the experiment, the search for a universal folding mechanism may provide sufficient benchmarks for theoretical models, at least for homology proteins. The proposed model is assumed to describe the first two steps in an event sequence oriented on reaching the native structure.

The early-stage structures presented in this paper were obtained as a result of a step back from the native structure of the protein. It resembles the frequently used procedure of polypeptide denaturation using molecular dynamics simulation (19, 23-26). However, a procedure allowing the Phi, Psi angle values (on the ellipse path) to be attributed to amino acids in proteins of unknown structure is under consideration and will be published soon.

Other support for the presented model is found in the suggestion that low-resolution structural features rather than high-resolution detail are more helpful in addressing the fundamental physics of the folding process (27).

The model presented in this paper is oriented only on testing the structure assumed to represent early stage folding. The influence of energy minimization and molecular dynamics simulation on the polypeptide structure changes was examined. In the result one can conclude, that both procedures produced reliable structural

forms. The unexpected structural motives did not appear. The comparison with experimental data is necessary (32-35). The simulation taking into account the influence of environment (ion strength, pH, ionization form of protein) is necessary and planed to be performed in the future. Particularly interesting is the comparison of results obtained based on H/^{2}H exchange experiments (34, 35) as well as of theoretical models (32, 33).

The common conclusion based on different experiments is that the secondary structure is rather preset in an early-stage folding process, what agrees quite well with the model presented here (35).

The early stage folding of lysozyme was analyzed experimentally using the hydrogen exchange labeling technique. The results revealed that the secondary structure is present in early stage folding although its localization can not be estimated. The helical fragments and some inter-helical links were found to be protected more rapidly. Although the irregular hinch fragment Trp63-Asn65 shows substantial inhibition of exchange due to contacts with helical domain, the β-sheet structural fragments gives no evidence for protection (34).

Our simulations appear to agree with these experimental results which is shown on non-bonding contacts map (Fig. 4D and Fig. 4F).

The firstly absent in ellipse-derived structure α-domain contacts, materialize after minimization procedure and are well maintained during the molecular dynamics simulation.

The segments constituting α-domain seem already to become organized (Fig. 2D, Fig. 4D), although next sequential steps are necessary to form the final native structure of the domain.

The significant role of the topology of the backbone and not of the particular side chains was found to dictate the initial dynamical behavior of a protein. It was supported by observations that different sequences having the same native topology fold in a similar manner. The observation was proved constructing the structures of protein chimeras (the sequence threading onto the backbone corresponding to the native state of a different protein of equal chain length) (33).

According to model presented in this paper, the folding pathway depends on the strategic points in polypeptide which is the location of prolines in a sequence.

The simulation of the second step of folding process expressing the hydrophobic interaction as responsible for higher packing will be adopted for the model in the way similar to those presented in (32).

References and Footnotes

1. D. O. V. Alonso, V. Daggett. *Prot. Sci. 7*, 860-874 (1998).
2. C. M. Dobson. *Phil. Trans. R. Soc. Lond. B 356*, 133-145 (2001).
3. A. Liwo, C. Czaplewski, J. Pillardy, H. A. Scheraga. *J. Chem. Phys. 115*, 2323-2347 (2001).
4. A. Liwo, P. Arfukowicz, C. Czaplewski, S. Odziej, J. Pillardy, H. A. Scheraga. *Proc. Natl. Acad. Sci. USA 99*, 1937-1942 (2002).
5. A. Fernàndez, K. Kostov, R. S. Berry. *Proc. Natl. Acad. Sci. USA 96*, 12991-12996 (1999).
6. A. Fernàndez, A. Colubri, G. Aqpigmanesi, T. Burastero. *Physica. A 293*, 358-384 (2001).
7. T. R. Sosnick, R. S. Berry, A. Colubri, A. Fernàndez. *Proteins Struct. Func. Gen. 49*, 15-23 (2002).

8. I. Roterman. *J. Theoretical Biol. 177*, 283-288 (1995).
9. I. Roterman. *Biochimie. 77*, 204-216 (1995).
10. W. Jurkowski, M. Bryliński, Z. Wiśniowski, L. Konieczny, I. Roterman. *Proteins Struct. Func. Bioinform 55*, 115-127 (2004).
11. S. Hayward. *Prot. Sci. 10*, 2219-2227 (2001).
12. R. L. Baldwin. *Science 295*, 1657-1658 (2002).
13. H. A. Scheraga. *Reviews in Computational Chemistry 3*, 73-142 (1992).
14. D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. DeBolt III, D. Ferguson, G. Seibel, P. A. Kollman. *Comp. Phys. Commun. 91*, 1-41 (1995).
15. J. Wang, P. Cieplak, P. A. Kollman. *J. Comput. Chem. 21*, 1049-1074 (2000).
16. D. Bashford, D. A. Case. *Annu. Rev. Phys. Chem. 51*, 129-152 (2000).
17. J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen. *J. Comput. Phys. 23*, 327-341 (1977).
18. C. A. Orengo, J. E. Bray, T. Hubbard, L. LoConte, I. Sillitoe. *Proteins Struct. Func. Gen. Suppl. 3*, 149-170 (1999).
19. A. R. Fersht, V. Daggett. *Cell 108*, 573-582 (2002).
20. M. Bryliński, W. Jurkowski, L. Konieczny, I. Roterman. *Bioinformatics 20*, 199-205 (2004).
21. N. Ferguson, A. R. Fersht. *Curr. Opin. Struct. Biol. 13*, 75-81 (2003).
22. J. Song, P. Bai, L. Luo, Z. Peng. *J. Mol. Biol. 280*, 167-174 (1998).
23. V. Daggett, M. Levitt. *J. Mol. Biol. 223*, 1121-1138 (1992).
24. V. Daggett, M. Levitt. *J. Mol. Biol. 232*, 600-619 (1993).
25. A. V. Finkelstein. *Protein Eng. 10*, 843-845 (1997).
26. C. L. Brooks III. *Curr. Opin. Struct. Biol. 8*, 222-226 (1998).
27. E. Alm, D. Baker. *Curr. Opin. Struct. Biol. 9*, 189-196 (1999).
28. C. M. Dobson, M. Karplus. *Curr. Opin. Struct. Biol. 9*, 92-101 (1999).
29. A. R. Clarke, J. P. Waltho. *Curr. Opin. Biotechn. 8*, 400-410 (1997).
30. K. Kuwajima, G. V. Semisotnov, A. V. Finkelstein, S. Sugai, O. B. Ptitsyn. *FEBS. Letters 334*, 265-268 (1993).
31. A. R. Dinner, A. Sali, L. J. Smith, C. M. Dobson, M. Karplus. *Trends Biochem. Sci. 25*, 331-339 (2000).
32. H. D. Nguyen, C. K. Hall. *Biotechnology and Bioengineering 80*, 823-834 (2002).
33. G. A. Arteca, K. Veluri. *Int. J. Quantum Chem. 88*, 220-225 (2002).
34. S. T. Gladwin, P. A. Evans. *Folding & Design 1*, 407-417 (1996).
35. O. Bieri, T. Kiefhaber. *J. Mol. Biol. 310*, 919-935 (2001).