



Retention of local conformational compactness in unfolding of barnase; Contribution of end-to-end interactions within quasi-modules

Kazuki Shinoda^{1,4}, Ken-ichi Takahashi², and Mitiko Go^{2,3}

¹*Division of Biological Science, Graduate School of Science, Nagoya University, Nagoya, Aichi 464-8602, Japan*

²*Department of Bioscience, Faculty of Bioscience, Nagahama Institute of Bio-Science and Technology, Nagahama, Shiga 526-0829, Japan*

³*Ochanomizu University, 2-1-1 Ohtsuka, Bunkyo-ku, Tokyo 112-8610, Japan*

⁴*Present address: Science & Technology Systems, Inc., 1-20-1 Shibuya, Shibuya-ku, Tokyo 150-0002, Japan*

Received 5 March, 2007; accepted 11 April, 2007

To understand how protein reduces the conformational space to be searched for the native structure, it is crucial to characterize ensembles of conformations on the way of folding processes, in particular ensembles of relatively long-range structures connecting between an extensively unfolded state and a state with a native-like overall chain topology. To analyze such intermediate conformations, we performed multiple unfolding molecular dynamics simulations of barnase at 498 K. Some short-range structures such as part of helix and turn were well sustained while most of the secondary structures and the hydrophobic cores were eventually lost, which is consistent with the results by other experimental and computational studies. The most important novel findings were persistence of long-range relatively compact substructures, which was captured by exploiting the concept of module. Module is originally introduced to describe the hierarchical structure of a globular protein in the native state. Modules are conceptually such relatively compact substructures that are resulted from partitioning the native structure of a globular protein completely into several contiguous segments with the least extended conformations. We applied this concept of module to detect a possible hierarchical structure of each snapshot structure in unfolding processes as well. Along with

this conceptual extension, such detected relatively compact substructures are named quasi-modules. We found almost perfect persistence of quasi-module boundaries that are positioned close to the native module boundaries throughout the unfolding trajectories. Relatively compact conformations of the quasi-modules seemed to be retained mainly by hydrophobic interactions formed between residues located at both terminal regions within each module. From these results, we propose a hypothesis that hierarchical folding with the early formation of quasi-modules effectively reduces search space for the native structure.

Key words: module, protein folding, molecular dynamics simulation, hierarchical folding, denatured state

In order to avoid Levinthal's paradox of protein folding, it is necessary for protein molecules to somehow reduce their conformational space to be searched for the native structures in the folding processes. The "new view" of protein folding describes this reduction of conformational space toward a native structure as a funnel-like shape of energy landscape along conformational degrees of freedom¹. How and what physical nature of proteins, then, reduces the conformational space? What does an ensemble of conformations in a process of reducing conformational space look like?

Corresponding author: Mitiko Go, Ochanomizu University, 2-1-1 Ohtsuka, Bunkyo-ku, Tokyo 112-8610, Japan, e-mail: gakucho@cc.ocha.ac.jp

Structural characterizations of transition states in several small proteins have been extensively progressed by the synergy of mutational phi-value analyses and molecular dynamics simulations²⁻⁷. Common features are seen that in transition states overall structures are very close to the native structures although their hydrophobic cores and secondary structures are considerably loosened. In the case of barnase, an intermediate state, which appears before the transition state, is also shown to have a native-like overall chain topology with substantial native-like contacts. How about earlier stages of folding?

The denatured form of staphylococcal nuclease was studied by NMR measurement of residual dipolar couplings (RDC) in weakly aligned state, and suggested to possess native-like long-range orderings of the chain segments even in 8 M urea solution⁸. However, it is recently reported that the RDC data are sufficiently explained with local structural preferences per residue without interpretation of long-range orderings⁹.

On the other hand, NMR relaxation measurements and chemical shift deviations indicates largely unstructured features with some short-range residual structures such as helix and turn, in denatured states of several proteins^{2-4,10-12}. Unfolding simulations also capture the denatured ensembles having the tendency for the corresponding short-range structures and some weak clustering of hydrophobic residues. In the case of unfolding simulation of barnase, a coarse-grained overall chain topology, although expanded, retains that of the native structure^{11,13}. If such a native-like overall topology is formed, the conformational space has already reduced to a substantial extent. How does a protein chain reach that topology from an earlier stage of folding (the earliest stage is a state of a newly born nascent chain)?

What ensembles of conformations connect between an extensively unfolded state with some short-range structures, if any, and a substantially ordered state with a native-like coarse-grained overall chain topology? To observe and characterize such ensembles of conformations before the formation of a native-like overall chain topology, we performed multiple unfolding molecular dynamics simulations of barnase at 498 K with explicit water. And we applied the concept of module to descriptively capture possible long-range features in unfolded conformations that have no native-like chain topology.

The concept of module is originally discovered in searching for a structural unit within globular domains that is related to the exon-intron structure of genes¹⁴. Module is defined by partitioning a globular domain into small contiguous segments that have the most compact or the least extended conformations¹⁴⁻¹⁶. The average length of modules is 13-residues long. Statistically significant correlations between module boundaries and intron positions are seen in various kinds of globular proteins¹⁴⁻¹⁸, which suggest that proteins have evolved by assembling modules through exon shuffling^{16,19}.

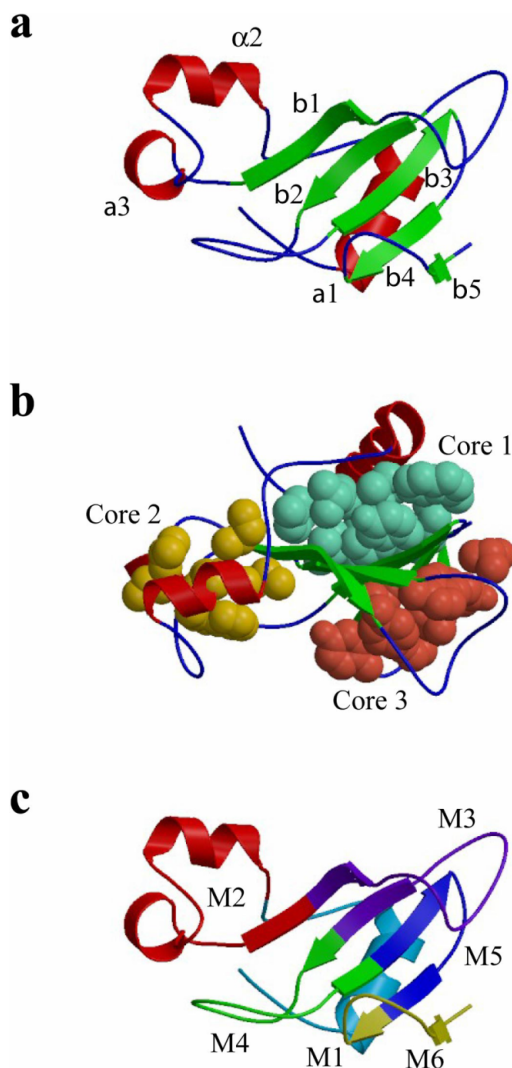


Figure 1 The three-dimensional (3D) structure and structural elements of barnase. (a) Secondary structures: α -helices and a β -sheet are shown in red and green, respectively. (b) hydrophobic cores: Core 1 (blue), core 2 (yellow) and core 3 (red) are shown by space-filling representation. (c) Modules: Six modules are shown in different colors; M1: sky blue, M2: red, M3: magenta, M4: green, M5: blue, and M6: yellow.

Quasi-independent features of modules are clarified, in particular by studying on modules of barnase (Fig. 1). Hydrogen bonds are mainly localized within modules²⁰. Five of the six isolated modules of barnase are mechanically stable, demonstrated by molecular dynamics (MD) simulations in vacuo and in explicit water²¹. These observations suggest that the native conformations of the modules are self-consistent, i.e., specified predominantly by intra-module interactions^{20,22}. Two modules of barnase, M2 and M3, each isolated in solution, were revealed by 2D NMR to have some secondary structures formed at their native positions²³. Mini-barnase, which lacks 26 amino acid residues closely corresponding to module M2 (residues 25-52), was revealed by CD and NMR spectroscopies to fold into a similar struc-

ture to those of barnase at least around the hydrophobic cores in a cooperative two-state manner^{24,25}. This measurement suggests a relatively independent relationship between module M2 and the rest of the molecule, mini-barnase. All these results propose the idea that modules would be quasi-independent folding units in a globular domain.

As mentioned above, module is originally introduced to describe a hierarchical structure of native protein globules. But the concept of modules is also applicable to non-native, flexible structures in a sense that it is possible in each snapshot conformation to find the way to divide the whole protein chain into relatively compact segments. Along with this conceptual extension, we hereafter call such detected relatively compact substructures “quasi-modules”. Using this description method, we can detect and characterize relatively long-range structures in unfolding simulations of barnase.

Materials and methods

Molecular dynamics simulations

All energy minimizations and MD simulations were performed with the sander module of AMBER 5 and sander_classic module of AMBER 6 program suites^{26,27}. We used amber 1994 (Cornell *et al.*) force field in all calculations²⁸. A crystal structure of barnase (PDB ID: 1RNB)²⁹ was utilized as the initial conformation of the simulations. The N-terminal first residue that is invisible in the crystal structure was omitted and the remaining 109 residues of barnase were used in our study. Assuming neutral pH condition, the N-terminal amino group and side chains of Lys and Arg were protonated, and the C-terminal carboxyl group and side chains of Asp and Glu were deprotonated. In this condition, net charge of barnase was +2. To relieve unfavorable atomic collisions, a 1000 step energy minimization was performed by the steepest descent method without cutoff truncation for non-bonded interactions. To neutralize the net charge of the whole system, seven sodium ions and nine chloride ions were set around barnase. Then, the molecular system was immersed in a cubic box containing TIP3P water molecules such that the minimum distance between any solute atom and the edge of the box was 15 angstrom. Number of water molecules was 8176. To equilibrate solvent molecules, a 20 ps MD simulation of water molecules and counter ions at 300 K at 1 atm under NPT condition was carried out. Then, a 1000 step energy minimization of the whole system was performed by the steepest descent method. This energy-minimized system was used as the initial conformation for the control simulation at 300 K. Further preparation was done for unfolding simulations at 498 K. The dimension of water box was isotropically expanded to adjust the density of the water to 0.829 g/ml, which corresponds to the lowest pressure (~26 atm) required for water to stay in the liquid state^{5,6,30}. Then, solvent molecules were re-equilibrated by a 20 ps MD simulation at

498 K under the constant volume condition, followed by a 1000 step energy minimization of the whole system by the steepest descent method. This energy-minimized system was used as the initial conformation for unfolding simulations at 498 K. Ten 3 ns MD simulations at 498 K under NVT condition were carried out as unfolding simulations of barnase. Periodic boundary condition and Particle mesh Ewald method were used to calculate electrostatic interactions without cutoff³¹. SHAKE algorithm was applied to fix covalent bond lengths involving hydrogen atoms to their equilibrium values³², and a time step of 2 fs was used in integrating the equation of motion. Snapshot structures were collected at every 0.5 ps for each simulation.

Identification of (quasi-)module boundaries

Modules are conceptually those sub-structural segments resulted from partitioning a globular domain into the most compact or least extended contiguous segments^{14,15}. Module boundaries are identified as residue positions that correspond to local minima in the centripetal profile^{16,20}. We have developed a full-automatic module identification method that uses only the coordinates of C^α-atoms of proteins (Go *et al.*, in preparation), which allows the method potentially applicable to any structure if the coordinate is available. Although the concept of module is originally given to describe the well-defined native structures of globular proteins as hierarchical structures, it can be applied to snapshot structures on the way of unfolding processes to describe the overall structure as the assembly of relatively compact sub-structural segments. We will use the term ‘quasi-module’ to distinguish this extended concept from the original one defined for the native structures. Using the same automatic method, module boundaries of the native structure and quasi-module boundaries of all snapshot structures in the unfolding trajectories were identified.

Definition of native contacts and Q-values

In this study, native contacts were defined by residue-based approach. We considered two residues to be in contact if the shortest distance between their constituent non-hydrogen atoms is less than 4.0 angstrom. Residues being in contact in a snapshot structure of MD simulations were considered to be in native contact if they were also in contact in the native structure. In counting the number of native contacts, only pairs of residues separated more than three residues apart from each other along the primary structure were considered. We introduced a progress variable Q, defined as a fraction of the number of residual native contacts for a certain conformation against the total number of native contacts in the native structure^{33–36}. This variable indicates the degree of unfolding; Q=1 corresponds to the native structure and Q=0 to fully denatured conformations without any native contacts. Native hydrophobic contacts are defined as native contacts between hydrophobic residues (Val, Ile, Leu, Met, Cys, Phe, Tyr, Trp). Barnase has three hydrophobic

cores³⁷. Persistency of each hydrophobic core is a fraction of residual native contacts in each hydrophobic core.

Results

Overall properties of unfolding trajectories

Trajectories of C^α-RMSD from the X-ray structure for the ten molecular dynamics simulations at 498 K show that barnase was progressively unfolded in various ways (Fig. 2). Conformations at 3 ns end points of the ten unfolding trajectories were largely different from the native structure and also different from each other (Fig. 3). Properties of these final structures are shown in Table 1. Seven of the ten unfolding trajectories reached C^α-RMSD of more than 10 angstrom. Less than 20% of the native hydrogen bonds were retained at 3 ns end points of 8 trajectories. These results demonstrate that various unfolding pathways of barnase were sampled in our simulations. Detailed analyses of persistence or loss of specific structural elements are described below.

Gradual losses of the native secondary structures

The native structure of barnase has three α -helices and a five-stranded anti-parallel β -sheet (Fig. 1a). Figure 4 shows the persistency of the native secondary structure elements (α -helices and β -sheet) along the time axis for each trajectory. Gradual losses of the native secondary structures, in particular, the β -sheet, were commonly seen although they proceeded in different manners among 10 trajectories. The α -helices were more sustainable than the β -sheet, the reason of which might be partly due to amber 1994 (Cornell *et al.*) force field²⁸, the force field we used, which is known as an α -helix-favoring force field. We next unified all the trajectories and analyzed them together to extract common properties among trajectories by introducing another progress variable Q, defined as a fraction of residual native contacts. As expected, Q values were gradually decreasing along time

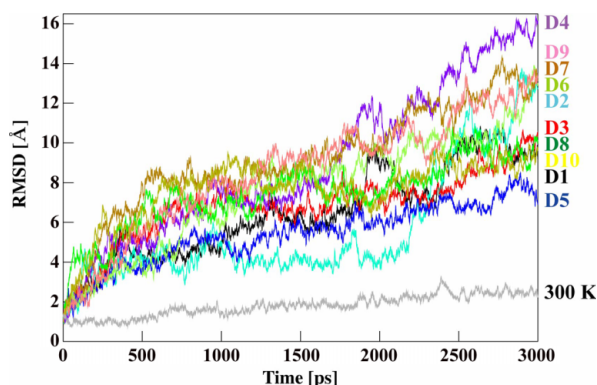


Figure 2 C^α-RMSDs from the X-ray structure as a function of time. C^α-RMSD trajectories of ten unfolding simulations (D1–D10) and of a control simulation at 300 K are shown. Each trajectory is indicated in different color.

axes in our unfolding simulations (data not shown). By classifying all snapshot structures based on their Q values, we can treat all trajectories together on the Q-axis. Following Tsai *et al.*³⁶, we classified all snapshot structures into four bins based on Q values; $1.0 \leq Q < 0.75$, $0.75 \leq Q < 0.5$,

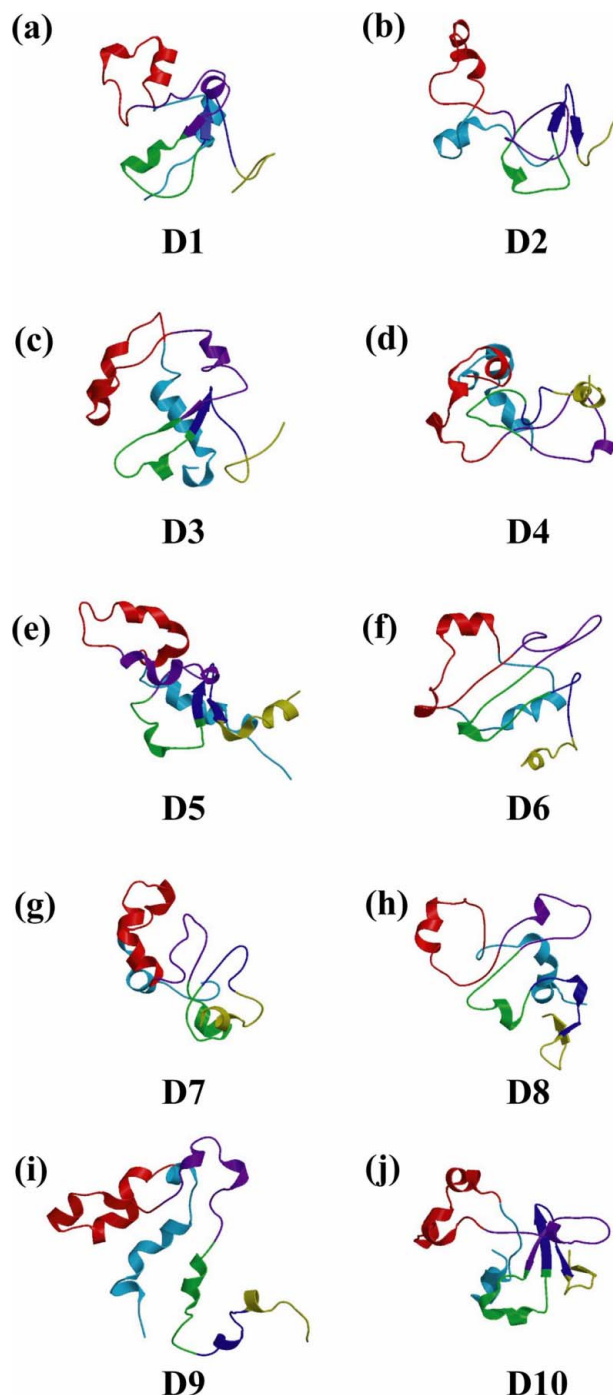
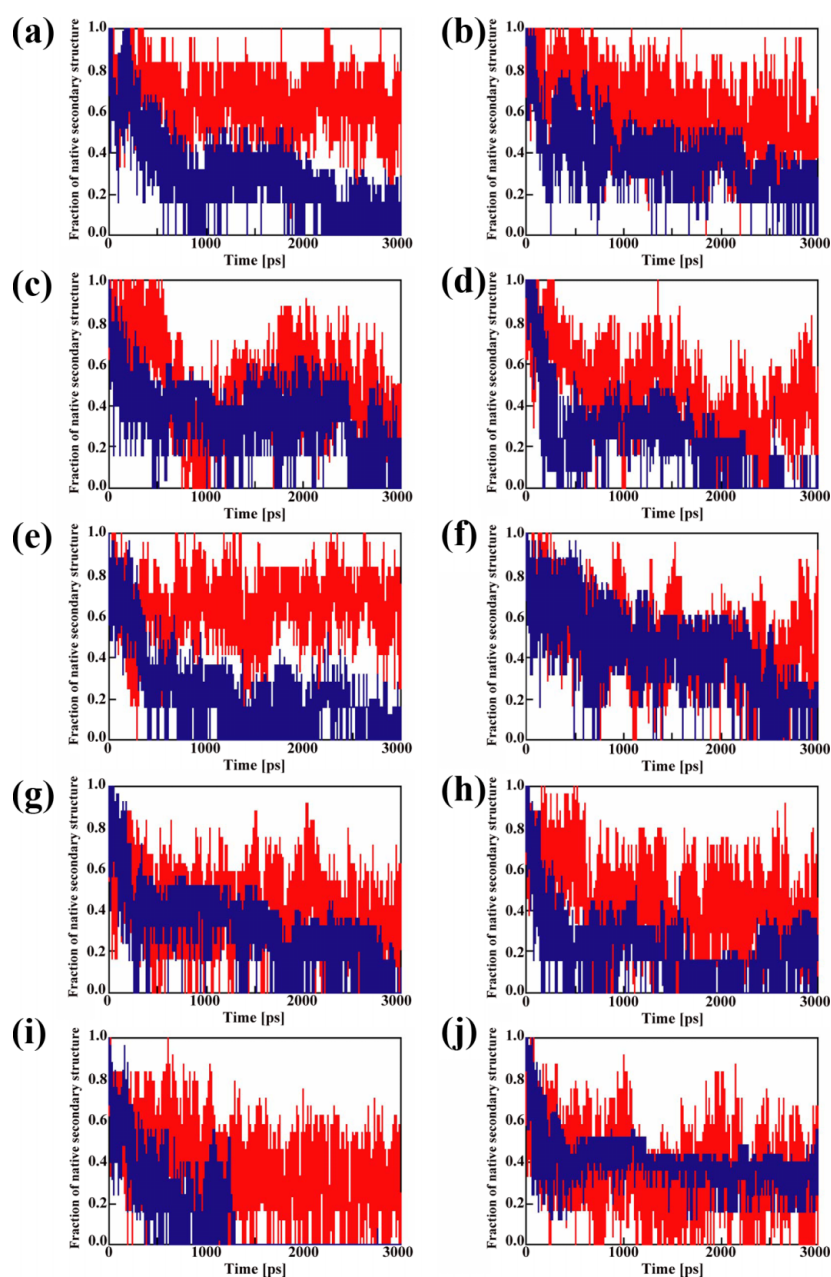


Figure 3 Final structures at 3 ns end points of ten unfolding trajectories. Secondary structures assigned by DSSP⁴² are shown by ribbon representation. Regions corresponding to native modules are shown in different colors. Color representation scheme of modules is the same as that of Figure 1 (c).

Table 1 Overall properties of the final structures in ten unfolding simulations (D1 to D10)

Properties	Native	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
RMSD (Å) ^a	–	9.2	12.8	10.3	16.0	6.9	13.0	13.2	10.1	15.4	9.6
Rg (Å) ^b	13.5	13.8	15.0	14.2	18.3	14.0	15.8	14.6	15.5	17.9	14.4
ASA (Å ²) ^c	5923	7548	8175	8098	9655	7711	8455	8127	8172	8585	7476
Q ^d	1.0	0.41	0.41	0.37	0.26	0.50	0.30	0.35	0.39	0.22	0.44
H-bond ^e	1.0	0.22	0.22	0.13	0.14	0.15	0.17	0.12	0.11	0.06	0.19

^a C^α-RMSD from the native structure.^b Radius of gyration.^c Accessible surface area.^d Q-value (a fraction of residual native contacts; see Materials and Methods).^e A fraction of residual native hydrogen bonds.**Figure 4** Persistency of the native secondary structures in each of ten unfolding simulations as a function of time. Trajectories of fractions of native α -helix and β -sheet are shown in red and blue, respectively. Secondary structures were assigned by DSSP⁴².

$0.5 \leq Q < 0.25$, and $0.25 \leq Q < 0.0$. These four bins were named, by taking medians, $Q=0.875$, $Q=0.625$, $Q=0.375$, and $Q=0.125$, respectively. We analyzed average persistency of the native secondary structures for each bin (Fig. 5a). In contrast to the time trajectories illustrated in Figure 4, Figure 5a clearly shows that both native α -helices and β -sheet were gradually lost as unfolding proceeded along the Q axis and that the α -helices were more stable than the β -sheet. At the most native-like ensemble (the bin of $Q=0.875$), both native secondary structure elements showed similar and high persistencies (α -helices: 0.729, β -sheet: 0.655), and the

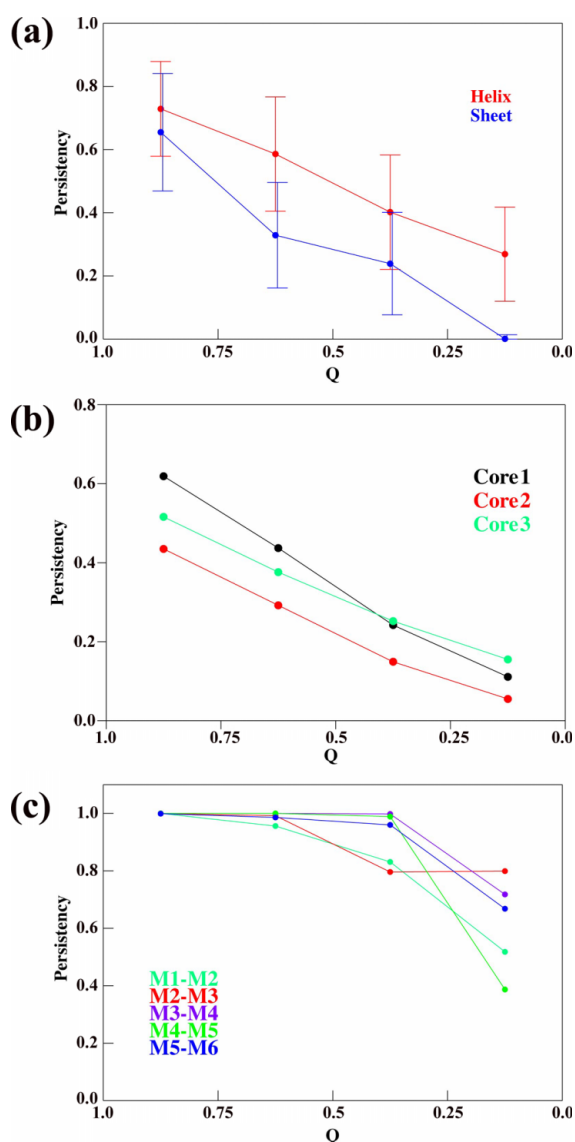


Figure 5 Average persistency of structural elements at each of the four unfolding bins. Persistency of native secondary structures (a), native contacts in hydrophobic cores (b) and quasi-module boundaries corresponding to the native module boundaries (c) is shown. The boundary between modules M1 and M2 is indicated as M1-M2; the other boundaries are indicated in the same way as M2-M3, M3-M4, M4-M5, and M5-M6.

persistencies gradually decreased as unfolding proceeded. Finally, the most unfolded ensemble (the bin of $Q=0.125$) has 0.269 of residual native α -helices and almost no native β -sheet (persistency: 0.001) (Fig. 5a). It should be noted that the regions (the C-terminal half of $\alpha 1$ and the turn connecting $\beta 3$ and $\beta 4$) which correspond to the residual local structures in urea-denatured barnase observed by an NMR relaxation measurement¹⁰, were found to retain their native-like conformations with high persistency even for the most unfolded ensemble (the bin of $Q=0.125$) in the simulations.

Gradual losses of the native hydrophobic cores

The native structure of barnase has three hydrophobic cores (Fig. 1b)³⁷. The major hydrophobic core (core 1) is formed between α -helix 1 and the β -sheet. The smallest hydrophobic core (core 2) is formed within a contiguous segments including $\alpha 2$, $\alpha 3$ and the N-terminal region of $\beta 1$. The other hydrophobic core (core 3) is formed between the β -sheet, the loop connecting $\beta 1$ and $\beta 2$, and the loop connecting $\beta 4$ and $\beta 5$. Using the same Q value as defined above for the analysis of secondary structures, we analyzed the persistency of hydrophobic cores in a similar way. We found that all three hydrophobic cores gradually broke as unfolding proceeded (Fig. 5b). At the initial stage of unfolding (the bin of $Q=0.875$), the three hydrophobic cores have already broken to a fair extent (persistency; core 1: 0.619, core 2: 0.435, and core 3: 0.516), and further disruptions proceeded up to the final stage of unfolding (the bin of $Q=0.125$), where almost no core structures retained (persistency; core 1: 0.111, core 2: 0.055, and core 3: 0.155). As described later, however, some native contacts between hydrophobic residues retained with high probability at the C-terminal region of $\alpha 1$, between $\beta 3$ and $\beta 4$ and so on.

Persistence of quasi-modules

The native 3D structure of barnase is decomposed into at least six modules, M1–M6; M1 (amino acid residues 1–24), M2 (25–52), M3 (53–73), M4 (74–88), M5 (89–98), and M6 (99–110) (Fig. 1c)²⁰. We decomposed all simulation snapshot conformations into quasi-modules by our automatic module-identification method. In the control simulation at 300 K, barnase retained its native structure throughout 3 ns of simulation time; the C^α -RMSD rose to only around 2 angstrom (Fig. 2). Therefore, the modules of barnase also kept their compact native conformations so that the module boundaries were completely retained in this control simulation (data not shown). In the case of the unfolding simulations, Figure 6 shows quasi-module boundaries along the time axis for each of 10 unfolding trajectories. The native module boundaries are indicated by pink bars with a width of 2 residues at both sides. The most surprising result of this analysis is that the positions corresponding to the native module boundaries were identified nearly perfectly as quasi-module boundaries throughout all the trajectories (Fig. 6). Only one exception is that the quasi-module boundary cor-

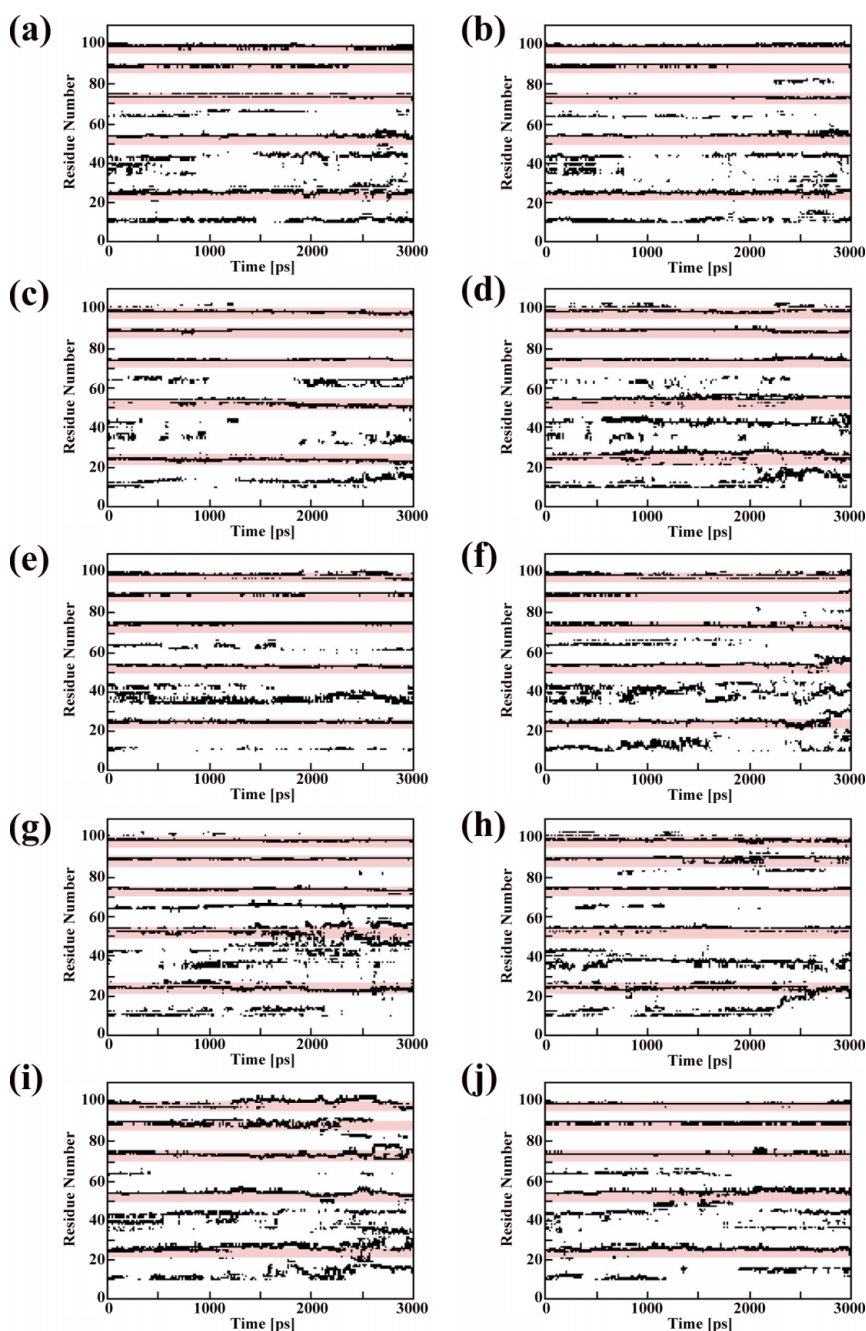


Figure 6 Comparison between native module boundaries and quasi-module boundaries (black dots) plotted as a function of time. Quasi-module boundaries were identified for each snapshot conformation in each trajectory of 10 unfolding simulations (a–j). Regions no more than 2 residues away from the native module boundaries are shaded in pink.

responding to the module boundary between modules M4 and M5 disappeared during last 500 ps of the trajectory D9 (Fig. 6i). On the other hand, there are many residues that are identified as quasi-module boundaries but not close to the native module boundaries. To examine whether these quasi-module boundaries away from the native module boundaries appear as often as those close to the native module boundaries, we plotted a frequency for each residue to be identified as a quasi-module boundary over all trajectories

(Fig. 7). From the plot, it is apparent that residues close to the native module boundaries have much higher frequencies to be identified as quasi-module boundaries than other residues.

We applied again the unifying analysis of all trajectories to see the retention of quasi-module boundaries corresponding to the native module boundaries (Fig. 5c). At bin $Q=0.875$, the boundary between modules M5 and M6 (M5-M6 boundary) was retained with a probability of 0.999

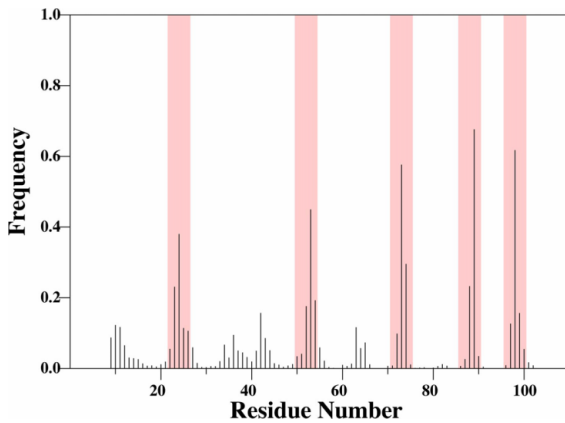


Figure 7 A frequency for each residue to be identified as a quasi-module boundary over all trajectories of unfolding simulations. Regions no more than 2 residues away from the native module boundaries are shaded in pink.

and all the other native boundaries were retained perfectly. At bin $Q=0.625$, all five native boundaries were nearly perfectly retained. At bin $Q=0.375$, boundaries M3-M4, M4-M5, and M5-M6 were still kept with very high persistency but retention probabilities of M1-M2 and M2-M3 boundaries decreased to about 0.8. Finally, even at bin $Q=0.125$, three boundaries M2-M3, M3-M4, and M5-M6 were highly retained (persistency; 0.80, 0.72, and 0.67, respectively). Retention probabilities of boundaries M1-M2 and M4-M5 decreased to 0.52 and 0.39, respectively. Average retention probability of all five boundaries at this most unfolded bin was 0.62. This is an unexpectedly high value because conformations at this ensemble contain less than 25% of all native contacts (the range of Q values at this bin is between 0 and 0.25) and persistencies of the native α -helix and the largest hydrophobic core were only 0.27 and 0.11, respectively.

Details of residual native contacts

To understand why the module boundaries of barnase were retained as quasi-module boundaries even if barnase was largely unfolded in the unfolding simulations, we examined where residual native contacts tended to be located. We first classified all residual native contacts into intra- and inter-module contacts, and compared their persistency. The number of residual native contacts was gradually decreased along the Q axis with barnase unfolding, but intra-module contacts were more retained than inter-module ones at all stages of unfolding (Fig. 8). Even at the most unfolded ensemble (the bin of $Q=0.125$), 37 of the intra-module native contacts on average were still retained, whereas the inter-module native contacts were almost disappeared (persistency: 0.04).

To further investigate detailed locations of residual native contacts, we drew contact maps where persistency of native contacts for each pair of residues are shown for each of the

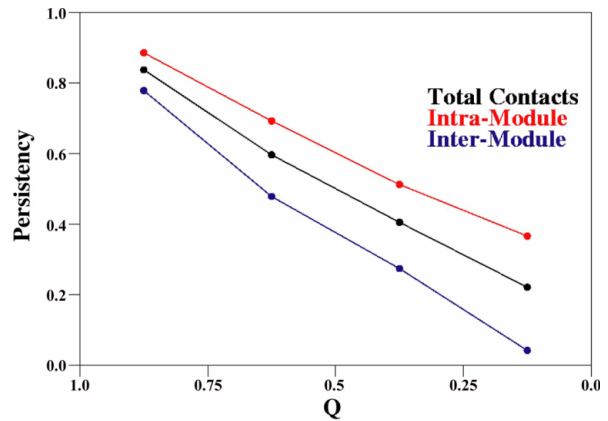


Figure 8 Average persistency of intra- and inter-module native contacts at each of the four unfolding bins.

ensembles in four different bins (Fig. 9 a–d (bottom left triangle halves)). As unfolding proceeded with the decrease of the Q value, persistency of native contacts were generally getting weaker. However, apart from high persistency of short-range contacts in helices, there is a tendency that intra-module contacts formed between terminal regions within each module are relatively well sustained compared to the inter-module contacts. Even at the most unfolded ensemble (the bin of $Q=0.125$), there exist clusters of the residual intra-module contacts between terminal regions for modules M2, M3, and M5.

We further focused on native contacts formed between only hydrophobic residues (Fig. 9 a–d top right triangle halves). We found relatively high persistency of hydrophobic contacts between terminal regions within each module except for the N-terminal module M1 and the C-terminal module M6. Figure 10 shows representative conformations of segments corresponding to modules M3 and M5, which are picked up from the most unfolded ensemble (the bin of $Q=0.125$). It seems that a cluster of hydrophobic residues formed between terminal regions within each module is a major factor to retain native-like compactness of modules.

Discussion

To elucidate a solution of the protein-folding problem, it is crucial to observe ensembles of relatively long-range structures that should connect between an extensively unfolded state and a state with a native-like overall chain topology. To observe such ensembles, we performed heat-induced unfolding simulations of barnase with an assumption that the unfolding process at high temperature significantly agrees with the reverse of folding process at physiological temperature. Although it is no doubt that free energy landscape of proteins inevitably changes when temperature of the system changes, Dinner and Karplus argued, from their lattice model simulations of protein folding/unfolding, that high temperature unfolding pathways resemble most closely the “fast

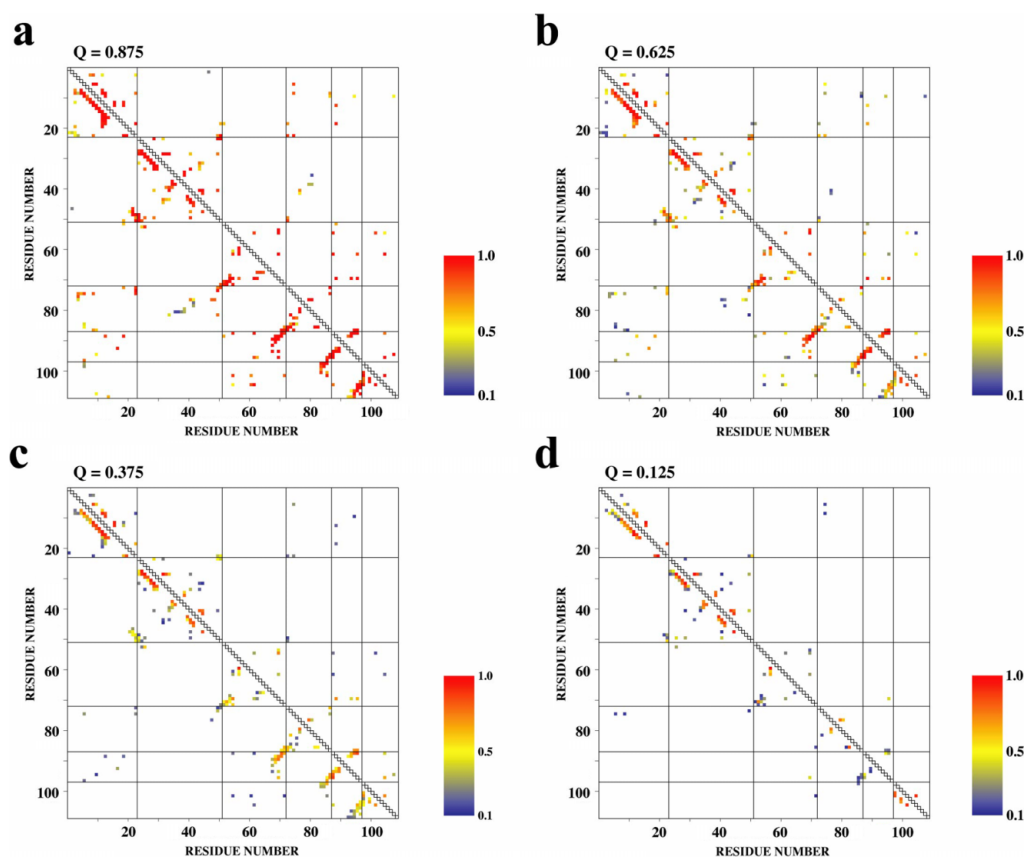


Figure 9 Inter-residue contact maps for each of the four unfolding bins (a: $Q=0.875$, b: $Q=0.625$, c: $Q=0.375$, d: $Q=0.125$). Average persistency of native contacts for each pair of residues is shown in color: gradation from red to yellow to blue with decreasing persistency between 1.0 and 0.1. All kinds of native contacts (bottom left triangle halves) and only native contacts between hydrophobic residues (top right triangle halves) are considered. The native module boundaries are delineated with horizontal and vertical lines. Small triangle areas on the diagonal lines correspond to intra-module regions.

track” of folding because high temperature makes the energy surface smooth³⁸. Our simulation results of largely unfolded states at high temperature were at least consistent with available experimental results of a denatured state at physiological temperature^{10,11} in the retention of residual local structures such as the C-terminal portion of $\alpha 1$ and the turn between $\beta 3$ and $\beta 4$, as described above. From these simulation data, we tried to extract a working-hypothesis concerning nascent long-range conformations in largely unfolded states, which are hard to be detected by experiments.

Our unfolding simulations of barnase were basically the same in the simulation protocol as that of Daggett’s group^{11,13} although using different software. In addition to the observation of the same residual local structures mentioned above, they observed persistence of roughly native-like chain topology in the unfolding simulations. On the other hand, more diverse conformations not necessarily possessing native-like chain topology were sampled in our simulations (see Fig. 3 and 11). This is probably because a larger number of independent unfolding simulations were performed in our study. Surprisingly, even in such more unfolded conformations, we observed with high probability some hierarchical

structure composed of quasi-modules (relatively long-range compact segments) corresponding in sequence position and compactness to the native modules (see Fig. 10 and 11). While Daggett’s group observed a segment of residues 25–55, which closely corresponds to module M2, behaving as a semiautonomous unit independent of the rest of barnase^{11,13}, we observed all of the segments corresponding to the native modules behaving relatively independent of each other in our simulations. Compactness of quasi-modules seems to be maintained by hydrophobic interactions between terminal regions within each module because relatively high persistency of native hydrophobic contacts was observed at such regions.

From these results, we propose a hypothesis that some hierarchical structure composed of quasi-modules appears in an early stage of protein folding, which effectively reduces search space for the native structure. Roughly speaking, native structure of module is a U-turn structure (see Fig. 10); both terminal parts come close to each other at the protein center and the middle part of a module goes out to protein periphery. And hydrophobic residues are generally often located near both terminal regions of modules (data not

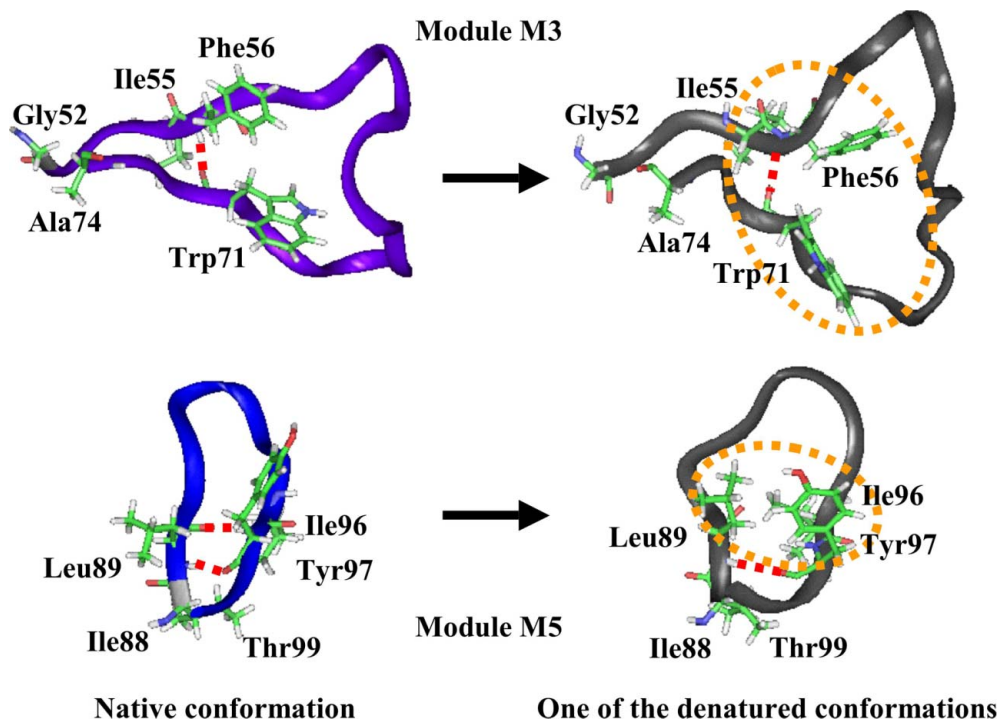


Figure 10 Representative conformations of segments corresponding to modules M3 and M5, which are picked up from the most unfolded ensemble (the bin of $Q=0.125$). The native structures of the modules are also shown for comparison. A hydrogen bond and a cluster of hydrophobic residues are shown by a red broken line and an orange broken circle, respectively.

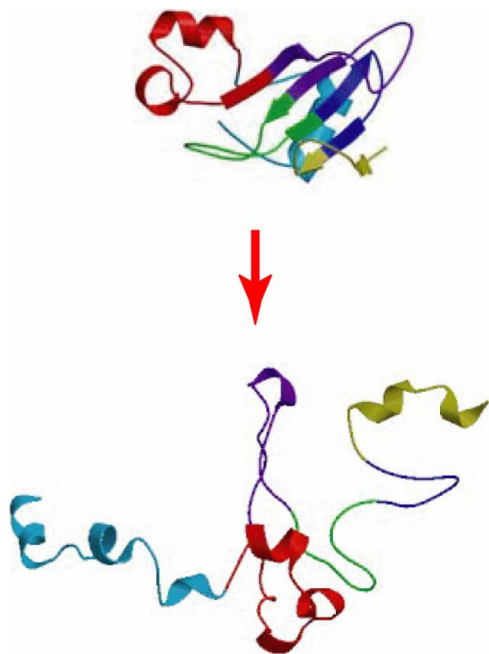


Figure 11 A representative unfolded structure of barnase. The structure at 3 ns end point of the trajectory D4 (bottom) is shown with the native structure of barnase (top). Color scheme is the same as that of Figure 1 (c).

shown). In an early stage of protein folding, hydrophobic collapse would occur more frequently between such terminal regions within each module than between those more distantly separated on the primary structure due to smaller loss of chain entropy in the former case than in the latter. These interactions draw both termini of each module close to each other, resulting in the acceleration of the formation of compact conformations of modules, and then followed by searching for native-like relative positioning of modules. It should be noted that Panchenko *et al.* found significant correlation between module and “foldon”, where the latter is defined as a kinetically competent, quasi-independently folding unit of a protein based on their energy landscape theory and identified as relatively the minimally frustrated contiguous segments at their native conformations as compared with their calculated molten-globule conformations^{39,40}. For barnase, two foldons (residues 1–30 and 31–55) were shown to fairly coincide with modules M1 (residues 1–24) and M2 (residues 25–52)³⁹. Correspondence between foldon and module in several proteins will strengthen the view that module folds rather independently in an early stage of protein folding.

So far, there seems to be no experimental evidence for formation of quasi-modules in denatured states of barnase. Some clue could be obtained from, for example, NMR experiments to measure long distances using site-directed spin and isotope labeling, which are successfully applied to determine roughly overall chain topology of barnase⁴¹.

Acknowledgments

Molecular dynamics calculations were done mainly in Information Technology Center of Nagoya University. This work was supported in part by Grants-in-Aid for Scientific Research (B) from Ministry of Education, Culture, Sports, Science and Technology of Japan to MG.

References

- Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nature Struct. Biol.* **4**, 10–20 (1997).
- Daggett, V. & Fersht, A. The present view of the mechanism of protein folding. *Nat. Rev. Mol. Cell Biol.* **4**, 497–502 (2003).
- Fersht, A. R. & Daggett, V. Protein folding and unfolding at atomic resolution. *Cell* **108**, 573–582 (2002).
- Daggett, V. Molecular dynamics simulations of the protein unfolding/folding reaction. *Acc. Chem. Res.* **35**, 422–429 (2002).
- Li, A. & Daggett, V. Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 by molecular dynamics simulations. *J. Mol. Biol.* **257**, 412–429 (1996).
- Li, A. & Daggett, V. Molecular dynamics simulation of the unfolding of barnase: Characterization of the major intermediate. *J. Mol. Biol.* **275**, 677–694 (1998).
- Daggett, V., Li, A., Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. Structure of the transition state for folding of a protein derived from experiment and simulation. *J. Mol. Biol.* **257**, 430–440 (1996).
- Shortle, D. & Ackerman, M. S. Persistence of native-like topology in a denatured protein in 8 M urea. *Science* **293**, 487–489 (2002).
- Bernado, P., Blanchard, L., Timmins, P., Marion, D., Ruigrok, R. W. & Blackledge, M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc. Natl. Acad. Sci. USA* **102**, 17002–17007 (2005).
- Freund, S. M. V., Wong, K. & Fersht, A. R. Initiation sites of protein folding by NMR analysis. *Proc. Natl. Acad. Sci. USA* **93**, 10600–10603 (1996).
- Wong, K.-B., Clarke, J., Bond, C. J., Neira, J. L., Freund, S. M. V., Fersht, A. R. & Daggett, V. Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. *J. Mol. Biol.* **296**, 1257–1282 (2000).
- Kazmirski, S. L., Wong, K. B., Freund, S. M., Tan, Y. J., Fersht, A. R. & Daggett, V. Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution. *Proc. Natl. Acad. Sci. USA* **98**, 4349–4354 (2001).
- Bond, C. J., Wong, K., Clarke, J., Fersht, A. R. & Daggett, V. Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: Description of the folding pathway. *Proc. Natl. Acad. Sci. USA* **94**, 13409–13413 (1997).
- Go, M. Correlation of DNA exonic regions with protein structural units in haemoglobin. *Nature* **291**, 90–92 (1981).
- Go, M. Modular structural units, exons, and function in chicken lysozyme. *Proc. Natl. Acad. Sci. USA* **80**, 1964–1968 (1983).
- Go, M. & Nosaka, M. Protein Architecture and the origin of introns. *Cold Spring Harbor Symp. Quant. Biol.* **52**, 915–924 (1987).
- Go, M. Protein structures and split genes. *Adv. Biophys.* **19**, 91–131 (1985).
- Sato, Y., Niimura, Y., Yura, K. & Go, M. Module-intron correlation and intron sliding in family F/10 xylanase genes. *Gene* **238**, 93–101 (1999).
- Gilbert W. Why genes in pieces? *Nature* **271**, 501 (1978).
- Noguti, T., Sakakibara, H. & Go, M. Localization of hydrogen bonds within modules in barnase. *Proteins: Struct. Funct. Genet.* **16**, 357–363 (1993).
- Takahashi, K., Oohashi, M., Noguti, T. & Go, M. Mechanical stability of compact modules of barnase. *FEBS Lett.* **405**, 47–54 (1997).
- Noguti, T. & Go, M. Modules of barnase: the physicochemical basis for their structures. in *Tracing Biological Evolution in Protein and Gene Structures* (Go, M. & Schimmel, P. eds), pp. 161–174, Elsevier, Amsterdam (1995).
- Ikura, T., Go, N., Kohda, D., Inagaki, F., Yanagawa, H., Kawabata, M., Kawabata, S., Iwanaga, S., Noguti, T. & Go, M. Secondary structural features of modules M2 and M3 of barnase in solution by NMR experiment and distance geometry calculation. *Proteins: Struct. Funct. Genet.* **16**, 341–356 (1993).
- Takahashi, K., Noguti, T., Hojo, H., Yamauchi, K., Kinoshita, M., Aimoto, S., Ohkubo, T. & Go, M. A mini-protein designed by removing a module from barnase: molecular modeling and NMR measurements of the conformation. *Protein Eng.* **12**, 673–680 (1999).
- Takahashi, K., Noguti, T., Hojo, H., Ohkubo, T. & Go, M. Conformational characterization of designed minibarnase. *Biopolymers* **58**, 260–267 (2001).
- Case, D. A., Pearlman, D. A., Caldwell, J. C., Cheatham III, T. E., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Ferguson, D. M., Radmer, R. J., Seibel, G. L., Singh, U. C., Weiner, P. K. & Kollman, P. A. *AMBER 5.0*, University of California, San Francisco (1997).
- Case, D. A., Pearlman, D. A., Caldwell, J. C., Cheatham III, T. E., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Radmer, R. J., Duan, Y., Pitner, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K. & Kollman, P. A. *AMBER 6*, University of California, San Francisco (1999).
- Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M. Jr., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
- Baudet, S. and Janin, J. Crystal structure of a barnase-d(GpC) complex at 1.9 Å resolution. *J. Mol. Biol.* **219**, 123–132 (1991).
- Haar, L., Gallagher, J. S. & Kell, G. S. *NBS/NRC Steam Tables: Thermodynamic and Transport Properties and Computer Programs for Vapor and Liquid States of Water in SI units*, Hemisphere Publication Corporation, Washington, D.C. (1984).
- Darden, T., York, D. & Pedersen, L. Particle mesh Ewald—an N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
- Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
- Sali, A., Shakhnovich, E. & Karplus, M. Kinetics of protein folding. A lattice model study of the requirements for folding to the native state. *J. Mol. Biol.* **235**, 1614–1636 (1994a).
- Sali, A., Shakhnovich, E. & Karplus, M. How does a protein fold? *Nature* **369**, 248–251 (1994b).

35. Lazaridis, T. & Karplus, M. "New view" of protein folding reconciled with the old through multiple unfolding simulations. *Science* **278**, 1928–1931 (1997).
36. Tsai, J., Levitt, M. & Baker, D. Hierarchy of structure loss in MD simulations of src SH3 domain unfolding. *J. Mol. Biol.* **291**, 215–225 (1999).
37. Serrano, L., Kellis, J. T., Jr., Cann, P., Matouschek, A. & Fersht, A. R. The folding of an enzyme. II. Substructure of barnase and the contribution of different interactions to protein stability. *J. Mol. Biol.* **224**, 783–804 (1992a).
38. Dinner, A. R. & Karplus, M. Is protein unfolding the reverse of protein folding? A lattice simulation analysis. *J. Mol. Biol.* **292**, 403–419 (1999).
39. Panchenko, A. R., Luthey-Schulten, Z. & Wolynes, P. G. Foldons, Protein structural modules, and exons. *Proc. Natl. Acad. Sci. USA* **93**, 2008–2013 (1996).
40. Panchenko, A. R., Luthey-Schulten, Z., Cole, R. & Wolynes, P. G. The foldon universe: a survey of structural similarity and self-recognition of independently folding units. *J. Mol. Biol.* **272**, 95–105 (1997).
41. Gaponenko, V., Howarth, J. W., Columbus, L., Gasmi-Seabrook, G., Yuan, J., Hubbell, W. L. & Rosevear, P. R. Protein global fold determination using site-directed spin and isotope labeling. *Protein Sci.* **9**, 302–309 (2000).
42. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).