Kodak Color Control Patches
Blue Cyan Green Yellow Red Magenta White 3/Color Black
© Kodak, 2007 TM: Kodak
Kodak Gray Scale
A 1 2 3 4 5 6 M 8 9 10 11 12 13 14 15 B 17 18 19
C Y M
© Kodak, 2007 TM: Kodak

Structural character of modules as evolutionary units and design of mini-barnase

（進化単位としてのモジュールの構造特性とミニ・バルナーゼの設計）

Ken-ichi Takahashi

高 橋 健 一

主論文

報告番号 甲第 3622 号

Structural character of modules as evolutionary units and design of mini-barnase

「進化単位としてのモジュールの構造特性と
ミニ・バルナーゼの設計」

Ken-ichi Takahashi
Laboratory of Molecular Biophysics, Department of Biology,
School of Science, Nagoya University
高橋健一
名古屋大学大学院理学研究科生物学専攻
動物学第III講座

CONTENTS

ABSTRACT

A globular domain of proteins is decomposed into modules. Modules are compact structural units consisting of contiguous 10 to 40 amino acid residues. Module boundaries are correlated with intron positions in genes encoding the proteins. This finding suggests that each module was once encoded by an exon and that exon shuffling produced various combinations of modules and resulted in creation of many different proteins.

If modules were original building blocks and were recruited into globular proteins, they could be mechanically stable. Such stability is required for useful parts to have their own conformation and/or function which are not influenced or deformed by surrounding partners. As module boundaries are often located on secondary structures, whether or not modules are mechanically stable elements like secondary structures has to be clarified.

To investigate the mechanical stability of modules, molecular dynamics simulations were performed on individually isolated modules of barnase. Barnase is a bacterial RNase from *Bacillus amyloliquefaciens*. Molecular dynamics study was performed for one nanosecond in vacuum and for 150 picoseconds in water. Five of the six excised modules of barnase retained native-like conformations during these simulations, showing they are mechanically stable. As a control study, similar simulations were carried out on pseudo modules; they were defined as segments starting at the center of one module and ending at the center of the following one. Contrary to modules, all pseudo modules were largely deformed in 100 picoseconds. The finding that the most modules investigated here have mechanical stability shows that modules are suitable as parts combined into proteins.

To determine which interaction mainly contributes to mechanical stability of modules, another series of molecular dynamics simulations were performed. The simulation was performed without hydrogen bonds and the interactions involving ionic charges. Four of the six modules of barnase retained their native-like conformations during this simulation. Thus, they are mechanically stabilized only by *van der Waals* and non-ionic partial charge interactions. Because these interactions are weak per atomic pair and only effective in a short distance, a sufficient number of atomic pairs must be in contact to accomplish the mechanical stability of modules. In fact the number of atomic pairs interacting in a module is not small because of its compact conformation. It is, therefore, concluded that the compactness of modules at the atomic level is essential for mechanical stability of modules.

Mechanical stability of modules has encouraged us to design a mini-barnase as a possible intermediate of evolving barnase. In the present study, a mini-barnase was designed by removing the second module M2 from barnase. The site-specifically ^{15}N-labeled mini-barnase was chemically synthesized and its conformation was examined by NMR measurement. Close correspondence in NMR chemical shifts between the mini-barnase and barnase was found. This result indicates that the mini-barnase folds into a stable conformation similar to that of barnase except for the removed module. It demonstrates that one module can be removed without altering structure of the other part in the parent protein and that modules are thus suitable building blocks as suggested from the molecular dynamics study of modules.

CHAPTER I
General Introduction

I-1. Protein modules

Positions of introns are correlated with boundaries of modules. Modules are compact structures of proteins. They consist of contiguous 10 to 40 amino acid residues [1-3]. This suggests that modules were once encoded by individual exons and that proteins were created through combination of modules by exon-shuffling.

I-2. Structural character of modules

Do modules have a structural character suitable for combination units? Such a character could be described by that modules retain their structures and/or functions when they are put into a new molecular environment by a new combination of modules. In other words, the structure of each module is specified by the amino acid sequence of the module itself. Without such an autonomy, a new combination of modules must result in only destruction of structure and function of the preceding combination. To what extent modules have an autonomy, is an important subject to be elucidated for understanding the structural basis of protein evolution as well as the principle of protein architecture.

To investigate the structural stability of isolated modules, I carried out

molecular dynamics simulations on individual modules of barnase, a bacterial RNase from *Bacillus amyloliquefaciens*, for one nanosecond in vacuum and for 150 picoseconds in water. Molecular dynamics simulation methods have been applied to biological problems such as studying the flexibility of proteins in ligand binding, determining protein structures from nuclear magnetic resonance (NMR) [4], evaluating molecular-modeled proteins in structural stability [5], studying the structural stability of peptide fragments in relation to protein folding [6]. However, the present work is the first application of molecular dynamics simulations to the problem of protein evolution, the structural stability of possible evolutionary units, modules.

Barnase is a single domain protein consisting of 110 amino acid residues and is decomposed into six modules M1 to M6. The size of the modules is in the range of 10 to 28 residues. As a control study, similar simulations were carried out on pseudo modules that were defined as segments starting at the center of one module and ending at the center of the following one. Five excised modules retained native-like conformations during the simulations while all pseudo modules were largely deformed. Hence, almost all modules of barnase are mechanically stable while pseudo modules are mechanically unstable.

I-3. Essential factor in mechanical stability of modules

I further examined the origin of the mechanical stability of modules by molecular dynamics simulations. Of the interatomic interactions within individual modules, nonbonded interactions consisting of electrostatic interactions, hydrogen bonds and *van der Waals* interactions stabilize module conformations. To investigate the role of these nonbonded interactions in the mechanical stability of modules, I simulated the isolated state of the modules by molecular dynamics calculation with hypothetically modified potential functions where hydrogen bonds and interactions involving ionic charges were

eliminated. It was found that *van der Waals* and non-ionic partial charge interactions were sufficiently strong for most modules of barnase to be mechanically stable. This result indicates that compact conformation of the modules is essential for their mechanical stability.

I-4. Design of mini-barnase

Molecular dynamics simulation shows that almost all modules of barnase have mechanical stability suitable for building blocks. This leads us to design a new protein by a new combination of modules. Such design will give us insight into protein evolution through various combinations of modules. In the present study, a mini-barnase was designed by removing the second module M2 from barnase and making a peptide bond between the flanking modules M1 and M3. Based on theoretical prediction method, two amino acid residues were replaced in order to increase the thermodynamic stability and solubility of the mini-barnase. The ^{15}N-labeled mini-barnase was chemically synthesized and its conformation was examined by NMR measurement. NMR is used for determination of the three dimensional (3D) structures of biological macromolecules in solution because parameters of NMR such as chemical shift, scalar coupling constant and nuclear Overhauser effect (NOE) contain structural informations [7]. Chemical shift is a sensitive probe detecting local difference in otherwise identical structures of related proteins, *e.g.*, wild type versus deletion mutant proteins [8]. If the 3D structure of the mini-barnase is similar to that of the corresponding portions of barnase, chemical shifts of respective residues will be also similar between the two proteins. Ten backbone amide nitrogens in the mini-barnase corresponding to those around one major hydrophobic core in barnase were site-specifically ^{15}N-labeled for readily detecting the chemical shifts of these sites. This hydrophobic core is important for overall fold of barnase. The NMR measurement revealed close correspondence in the chemical

shifts of ^{15}N-labeled amides between the mini-barnase and barnase. This result indicates that the structure of the M2-deleted mini-barnase is similar to that of intact barnase, at least for region of the important hydrophobic core.

I-5. References

1. Gō, M. Correlation of DNA exonic regions with protein structural units in haemoglobin. Nature 291:90-92, 1981.

2. Gō, M., Nosaka, M. Protein architecture and the origin of introns. Cold Spring Harb. Symp. Quant. Biol. 52:915-924, 1987.

3. de Souza, S.J., Long, M., Schoenbach, L., Roy, S.W., Gilbert, W. Intron positions correlate with module boundaries in ancient proteins. Proc. Natl. Acad. Sci. U.S.A. 93:14632-14636, 1996.

4. Karplus, M., Petsko, A. Molecular dynamics simulations in biology. Nature 347:631-639, 1990.

5. Du, P., Collins J.R., Loew, G.H. Homology modeling of a heme protein, lignin peroxidase, from the crystal structure of cytochrome c peroxidase. Protein Eng. 5:679-691, 1992.

6. van der Spoel, D., Vogel, H.J., Berendsen, H.J.C. Molecular dynamics simulations of N-terminal peptides from a nucleotide binding protein. Proteins 24:450-466, 1996.

7. Wüthrich, K. "NMR of proteins and nucleic acids." New York: Wiley, 1986.

8. Baldisseri, D.M., Torchia, D.A. Deletion of the Ω-loop in the active site of staphylococcal nuclease. 2. Effects on protein structure and dynamics. Biochemistry 30:3628-3633, 1991.

CHAPTER II
Mechanical Stability of Compact Modules of Barnase

Abstract: Globular proteins are composed of structural elements such as secondary structures and modules. Modules are compact segments consisting of 10 to 40 contiguous amino acid residues and are often encoded by exons. Therefore, the view that the modular organization of proteins is a result of exon-shuffling or -fusion is given support. Secondary structures such as α-helix and β-sheet are stabilized by hydrogen bonds and are thus considered to be stable, structural elements of a globular domain. Since module boundaries are often located on α-helices or β-sheets, it is not obvious whether the modules are mechanically stable. We carried out molecular dynamics simulations on modules of barnase, a bacterial RNase from *Bacillus amyloliquefaciens*, for one nanosecond *in vacuo* and 150 picoseconds in water. Five of six modules (M1:1-24, M2:25-52, M3:53-73, M4:74-88, M5:89-98) retained native-like conformations during these simulations. Only the C-terminal module (M6:99-110) was deformed; it is less compact than the other modules. As the modules are mechanically stable they are suitable as parts combined into proteins. Together with RNase activity of the three isolated modules of barnase, M2, M3 and M6, our study supports the view that modules were indeed original

building blocks of proteins.

II-1. Introduction

Introns are non-coding DNA sequences that split various genes encoding proteins. In some genes, intron positions are closely correlated with structural elements of the proteins gene-encoded [1-11]. This means that such structural elements were the original building blocks and were recruited into proteins by exon-shuffling [12,13]. There are two contradicting hypotheses on the origin of introns. One is the "intron early" [6,12,14] and the other is the "intron late" [15]. The finding of a close correlation of intron positions with protein structural units supports the intron early hypothesis.

Modules, candidates of such original building blocks of proteins, are compact substructures within a globular domain [2,3]. Correlation between module boundaries and intron positions was found in hemoglobin, lysozyme, ovomucoid 3rd domain and triose phosphate isomerase [1-8].

If modules are original building blocks and are recruited into globular proteins, they could be mechanically stable. Such stability is required for useful parts to have their own conformation and/or function which are not influenced or deformed by surrounding partners. However, other investigators reported that no correlation was observed between introns and secondary structures [15,16]. As module boundaries are often located on secondary structures, whether or not modules are mechanically stable elements like secondary structures has to be clarified. Since module boundaries show a close correlation with intron positions ([10] and unpublished results), mechanical stability must be a crucial intrinsic character of modules, as original building blocks of proteins. We studied the molecular dynamics of barnase, a small enzyme.

Barnase, a bacterial RNase from *Bacillus amyloliquefaciens* [17], consists of 110 amino acid residues and has no disulfide bond. The three-

dimensional structure (3D) of barnase was determined by X-ray crystallography [18-20] and NMR spectroscopy [21]. Barnase was decomposed to six modules, M1 to M6 [22]. Three of the six modules were found to have weak RNase activity [23], suggesting that they were primitive mini-enzymes, as independent segments.

The six compact modules are connected spatially to form the globular structure of barnase. In intact barnase, each module interacts with other modules and solvent. If these interactions are eliminated, the module may be deformed, or alternatively it may retain native-like conformations. This hypothetical state of the isolated module *in vacuo* can be simulated by molecular dynamics. Many mechanical properties of a protein molecule, such as conformational changes caused by ligand binding or by amino acid substitution, are characterized by dynamics in the time range of pico to nano seconds. We carried out molecular dynamics simulations of the six isolated modules for a nanosecond *in vacuo* starting from their native conformations. If a module retains conformation similar to its intact one during the simulation, the 3D structure of the module is mechanically stable. The mechanical stabilities of "pseudo modules" were also studied by molecular dynamics, as a control study. Pseudo modules were defined as segments starting at the center of one module and ending at the center of the following one [22]. Five pseudo modules, P1 to P5 were defined as P1:13-38, P2:39-62, P3:63-80, P4:81-93 and P5:94-104. Molecular dynamics simulations were carried out for the pseudo modules isolated *in vacuo* for fifty picoseconds. Our study was extended to simulation of modules and pseudo modules, in explicit inclusion of water molecules as solvent. We then compared the mechanical stability of modules *in vacuo* and in water.

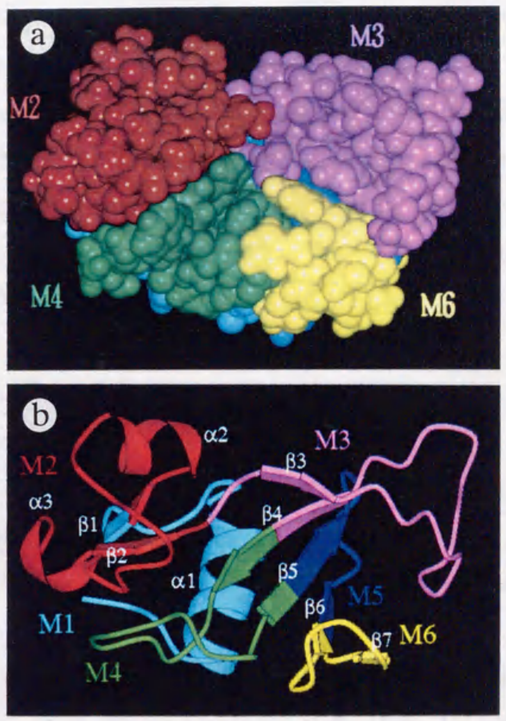

Fig. II-1. Module structure of barnase. The 3D structure of barnase [18] is shown from the direction of the catalytic site. Modules M1 to M6 are colored differently; M1 in sky blue, M2 red, M3 magenta, M4 green, M5 blue and M6 yellow. (a) A space-filling model. (b) A ribbon model of the mainchain drawn with the program MolScript [32]. Three α-helices, α1 to α3, and seven β-strands, β1 to β7, were identified by the algorithms of Kabsch and Sander [33].

II-2. Materials and methods

II-2.1. Modules and the secondary structures

Module boundaries were identified by a computerized method as 24, 52, 73, 88 and 98 [22]. Thus, the six modules M1 to M6 were identified as 1-24, 25-52, 53-73, 74-88 and 99-110 (Fig. II-1). Barnase has three α-helices,α1 to α3, (7-17, 27-32, 42-45), and seven β-strands, β1 to β7 (24-25, 50-51, 55-56, 71-

75, 87-91, 96-99, 107-108) (Fig. II-1 b). The first and second β-strands form a parallel β-sheet and the other five form an anti-parallel β-sheet. Helix α1 belongs to module M1. Helices α2 and α3, and the parallel β-sheet are involved in module M2. Module boundaries between M1 and M2, between M3 and M4, between M4 and M5, and between M5 and M6 are located on parallel or anti-parallel β-sheets (Fig. II-1 b). The anti-parallel β-sheet is in the four consecutive modules M3 to M6.

II-2.2. *In vacuo* simulation

Conformations of fragments corresponding to modules or pseudo modules of barnase [18] were taken as the initial structures for simulations. To modules M1 to M6, we assigned amino acid residues 1 to 24, 25 to 52, 53 to 73, 74 to 88, 89 to 98, and 99 to 110, respectively [22]. We used no overlapping residues. To pseudo modules P1 to P5, we assigned residues 13 to 38, 39 to 62, 63 to 80, 81 to 93, and 94 to 104, respectively. These five pseudo modules correspond to pseudo modules m2 to m6, respectively, defined by Noguti, *et al.* [22]; the other pseudo modules m1 (3-12) and m7 (105-110) were not included in this study, because they consist of only half part of a module and thus they cannot be compared with modules regarding conformational stability. Each polypeptide fragment was terminated at the peptide bond, without charged nor capped ends. Lys, Arg, Asp and Glu were ionized. Each module or pseudo module was treated as being isolated *in vacuo*, except that a linear distance dependent dielectric function, ε=r, was applied. Non-bonded interactions were included without cutoff. Each fragment was energy minimized until the root mean square (RMS) gradient of the potential was less than 0.01 kcal/mol/Å in the cycles of conjugate-gradient. We found that the use of 0.1 kcal/mol/Å instead of 0.01 kcal/mol/Å as the criterion of convergence of energy minimization did not alter substantially the results of the following molecular dynamics simulations.

Energy-minimized conformations were used as initial conformations of the molecular dynamics simulations. Molecular dynamics of individual modules and pseudo modules were carried out for a nanosecond and 50 picoseconds, respectively. The time steps were 0.25 femtosecond. The system temperature in the molecular dynamics was kept at 300 K by the heat-bath coupling method [24] with coupling constant 0.2 picosecond. Snapshot structures per picosecond in each simulation were stored as trajectory and averaged over the last half-trajectory to obtain the averaged simulation structure. The AMBER (rev.3a) energy parameters of united-atom model and programs [25] were used. We also carried out the same series of simulations using DREIDING energy parameters [26], the objective being to confirm the results obtained using AMBER energy parameters.

II-2.3. Solution simulation

Fragments corresponding to modules and pseudo modules were extracted from the X-ray crystal structure atomic coordinate of barnase [20] (BNL code: 1bgs) as initial structures for simulations. We used overlapping fragments consisting of residues 1 to 24, 24 to 52, 52 to 73, 73 to 88, 88 to 98, and 98 to 110, as modules M1 to M6, respectively in the solution simulation. Overlapping of residues at the terminals was used in water simulation so as to be consistent with previous experimental studies [23,27]. Pseudo modules, P1 to P5 consist of 12 to 38, 38 to 62, 62 to 80, 80 to 93, and 93 to 104, respectively. Both terminals of each polypeptide fragment were charged. Lys, Arg, Asp and Glu were ionized. Each fragment was energy minimized *in vacuo* with ε=r and no cut-off until the RMS gradient of potential was less than 0.1 kcal/mol/Å. The energy-minimized structures were spherically surrounded by water molecules such that width of the water shell exceeded 12 Å. Radii of water spheres were 28 Å for modules and 29 to 36 Å for pseudo modules. Numbers of water molecules were

about 3000 for modules and 3000 to 6000 for pseudo modules. Water molecules were restrained within the spheres by a soft half-harmonic potential with a force constant of 1.5. The dielectric constant ε=1 and cut-off of 12 Å were used for minimization. The dielectric constant ε=1, cut-off of 9 Å, and time steps of 1 femtosecond were used for molecular dynamics. The temperature in the molecular dynamics was initially 0.1 K and was kept at 300 K by the heat-bath coupling method [24] with a coupling constant of 0.2 picosecond for solute and 0.4 for solvent. With solutes constrained, water molecules were energy minimized until the RMS gradient of potential was less than 1 kcal/mol/Å, and then subjected to 20 picosecond molecular dynamics. All atoms of the systems were energy minimized until the RMS gradient of potential was less than 1 kcal/mol/Å. The resultant structures were used for the initial conformations of 150 picosecond molecular dynamics simulations of modules and pseudo modules in water. Snapshot structures per picosecond in each simulation were stored and averaged over the last 50 picosecond trajectory to obtain the averaged simulation structure. The AMBER (rev.3a and 4) energy parameters of all-atom model and programs [28] were used.

II-2.4. Conserved hydrogen bonds

A hydrogen bond formation was defined under the following conditions; the distance between the hydrogen atom and the acceptor atom was less than 2.5 Å, and the angle between the chemical bond from the donor to the hydrogen atom and the hydrogen bond from the hydrogen atom to the acceptor exceeded 120 degrees. For *in vacuo* simulation, we defined conserved hydrogen bonds during the simulation of each module as the hydrogen bonds formed in more than 30% of conformations in the last 500 picosecond trajectory of the simulation. For solution simulation, the conserved hydrogen bonds are defined as those formed in more than 80% of conformations in the last 50 picosecond

trajectory.

II-3. Results

II-3.1. *In vacuo*

Five modules M1 to M5, but not the C-terminal module M6, maintained native-like conformations through one nanosecond simulations *in vacuo* (Fig. II-2 a-e, Table II-1). Relaxations of potential energy in the range of 40 to 115 kcal/mol were observed in the first 500 picoseconds, and systematic relaxations were not observed in the last 500 picoseconds, in all modules. This shows that the modules reached mechanically stable conformations in the first 500 picoseconds. Root mean square deviation (RMSD) of C^{α} atom positions between X-ray structures of modules, M1 to M5, and the average conformations in the last 500 picosecond trajectories of the simulations were in the range of 2.3 to 4.4 Å (Table II-1). Mainchains in the average conformations of modules M1 to M5 maintained compact conformations, similar to their native ones (Fig. II-2 a-e).

Hydrogen bonds important for specifying the native-like conformations of modules M1 to M5 were conserved during the simulations (Fig. II-2 a-e, Table

Table II-1. Root mean square deviations (RMSD) of the simulated conformations of modules and pseudo modules *in vacuo* from the X-ray structures.*

Module	RMSD (Å)	Pseudo module	RMSD (Å)
M1 (1 - 24)	2.3	P1 (13 - 38)	8.2
M2 (25 - 52)	4.4	P2 (39 - 62)	8.3
M3 (53 - 73)	3.4	P3 (63 - 80)	8.9
M4 (74 - 88)	3.0	P4 (81 - 93)	8.5
M5 (89 - 98)	2.6	P5 (94 -104)	5.0
M6 (99 -110)	4.7		

*Conformations in the trajectories were averaged in the last 500 (25) picoseconds of the one nano (50 pico) second molecular dynamics of each module (pseudo module). N- and C-terminal residue numbers of modules (pseudo modules) are given in the parentheses. The RMSD of C^{α} atom positions of each module (pseudo module) was calculated between the resultant average conformation and the corresponding X-ray structure [18].

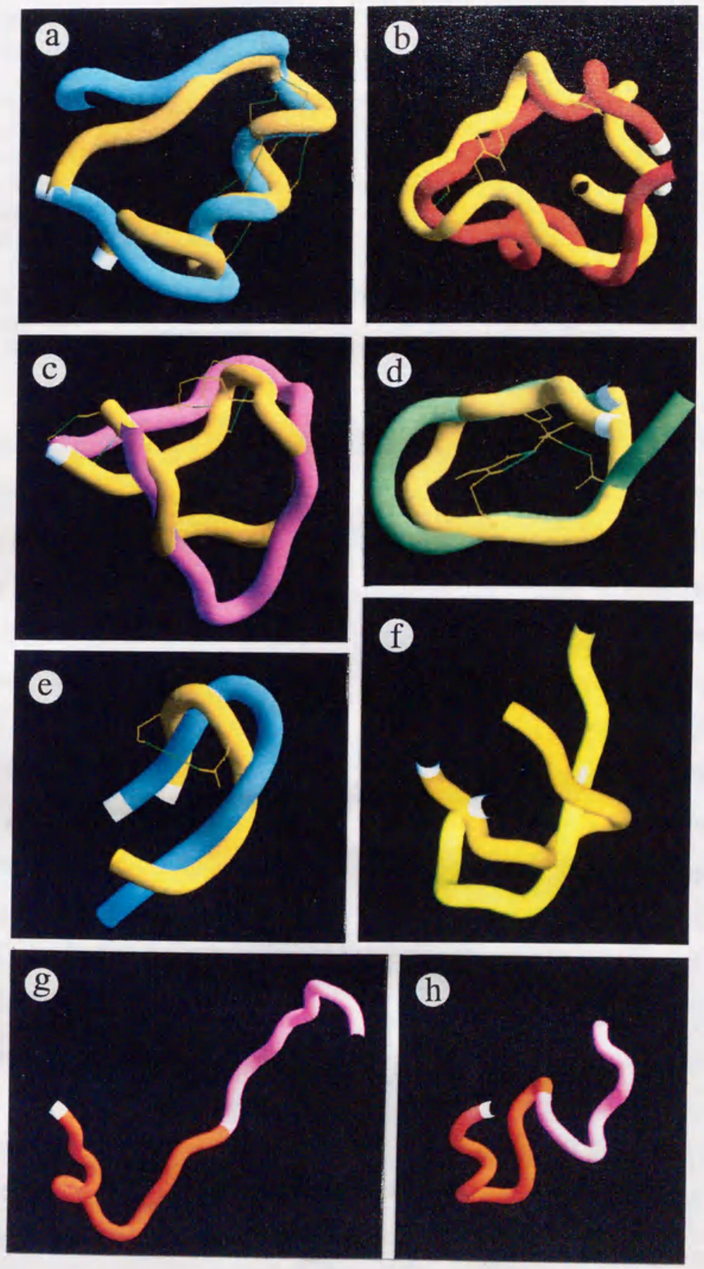

Fig. II-2. (Legend see following page).

Fig. II-2. The averaged simulation structures of modules and a pseudo module *in vacuo* and their corresponding X-ray crystal structures. Conformations were averaged over the trajectories in the last 500 picoseconds of the one nanosecond molecular dynamics of each module. The resultant average conformation was superimposed onto the corresponding part of the X-ray structure of barnase [18]. The mainchain folds are shown in tube models. X-ray structures of modules are in the same color as for Fig. II-1. The N-terminus of each fragment is indicated in white. The simulated average conformation of a module is in dark yellow. Modules, M1 to M6, are shown in **a** to **f,** respectively**.** The conserved hydrogen bonds (see methods) are shown in green lines together with the related sidechain and mainchain. **g**, The X-ray structure of pseudo modules P2, consisting of the C-terminal half of M2 (in red) and the N-terminal half of M3 (in magenta). **h**, The average conformation of pseudo module P2, calculated from the last 25 picosecond trajectory of the 50 picosecond dynamics. (See figure on preceding page.)

II-2). Particularly hydrogen bonds in turns and/or between the N- and C-termini of each module, which are important to maintain compactness, were conserved. Hydrogen bonds Lys39 N-Val36 O, Gly40 N-Ala37 O and Asn41 N^{δ}-Gly34 O stabilizing the turn consisting of residues 36-40 in the center of M2 were observed in more than 88% conformations in the last 500 picosecond trajectory of the simulation. The N- and C-termini of M3 form a part of the anti-parallel β-sheet in the native conformation. Two hydrogen bonds in this part of the β-sheet, Gly53 N-Glu73 O and Glu73 N-Asp54 O, were kept in about 40% of the simulated conformations. The hydrogen bond between sidechains of Glu60 and Lys62 stabilizing a β-turn of M3 was conserved completely. The compact form of M4 was maintained by three conserved hydrogen bonds Asn77 N^{δ}-Asp86 O^{δ}, Arg83 N^{θ}-Asp75 $O^{\delta 1}$ and Arg83 N^{ε}-Asp75 $O^{\delta 2}$; they contributed to maintain the N- and C-terminal halves of M4 in close proximity. The β-turn consisting of residues 91-94 in the center of M5 was stabilized by the hydrogen bond Ser91 O^{γ}-Asp93 O^{δ}, which was formed in 32% of the simulated conformations. To maintain the compact conformation of M1, helix α1 must be retained. The hydrogen bonds in helix α1 were formed in 35% to 99% conformations in the last 500 picosecond trajectory of the simulation.

Secondary structures observed in the native conformations of the modules

Table II-2. Conservation of hydrogen bonds in the molecular dynamics (MD) simulations of modules.

Hydrogen bonds observed in the X-ray crystal structure	Percentage of time ensemble in which each hydrogen bond exists	
	MD *in vacuo* (500-1000ps)	MD in water (100-150ps)
Module M1:		
between both termini:		
Asn23 N^{δ}....O Ile4	0	0
helix α1:		
Val10 N.....O Thr6	0	* 84
Ala11 N.....O Phe7	* 92	* 96
Asp12 N.....O Asp8	1	*100
Tyr13 N.....O Gly9	* 56	* 94
Leu14 N.....O Val10	* 96	* 98
Gln15 N.....O Ala11	* 91	* 98
Thr16 N.....O Asp12	* 99	* 98
Thr16 O^{γ}....O Asp12	0	* 94
Tyr17 N.....O Tyr13	* 35	*100
His18 N.....O Leu14	* 43	* 94
Module M2:		
between both termini (a part of β-sheet):		
Gly52 N.....O Ile25	0	0
Ile25 N.....O Ser50	1	* 90
turn in the center:		
Val36 N....O^{δ} Asn41	6	74
Lys39 N.....O Val36	* 88	10
Gly40 N.....O Ala37	* 95	0
Asn41 N^{γ}....O Gly34	* 99	36
helix α2:		
Ala30 N.....O Thr26	0	* 94
Gln31 N.....O Lys27	* 96	* 96
Ala32 N.....O Ser28	1	* 94
Leu33 N.....O Glu29	0	* 96
Leu33 N.....O Ala30	* 96	* 96
helix α3:		
Asp44 N.....O Asn41	0	56
Val45 N.....O Asn41	0	* 86
Ala46 N.....O Leu42	0	* 96

Table II-2. (continued)

Hydrogen bonds observed in the X-ray crystal structure	Percentage of time ensemble in which each hydrogen bond exists	
	MD *in vacuo* (500-1000ps)	MD in water (100-150ps)
Module M2:		
others:		
Thr26 N...$O^{\varepsilon 2}$ Glu29	80	96
Glu29 N....O^{γ} Thr26	1	62
Gly34 N.....O Gln31	1	26
Trp35 N.....O Ala30	0	78
Lys49 N.....O Ala46	2	30
Module M3:		
between both termini (a part of β-sheet):		
Glu73 N.....O Asp54	* 41	0
Gly53 N.....O Glu73	* 34	0
Phe56 N.....O Trp71	0	0
turn 1:		
Asn58 N^{δ}....O Leu63	0	0
Gly61 N.....O Asn58	* 44	0
Lys62 N^{ζ}...$O^{\varepsilon 1}$ Glu60	*100	0
Leu63 N....O^{δ} Asn58	0	0
turn 2:		
Arg69 $N^{\eta 1}$...O Pro64	0	0
Trp71 N^{ε}....O Pro64	0	0
Module M4:		
between both termini (including a part of β-sheet):		
Ile88 N.....O Ala74	0	0
Ala74 N.....O Ile88	0	62
Ile76 N.....O Asp86	0	0
Asn77 N^{δ}...$O^{\delta 2}$ Asp86	* 96	0
Arg83 N^{ε}...$O^{\delta 2}$ Asp75	* 50	0
Arg83 $N^{\eta 2}$..$O^{\delta 1}$ Asp75	* 87	* 86
Asn84 N^{δ}....O Ile76	0	0
others:		
Asp86 N....O^{δ} Asn84	0	42
Arg87 $N^{\eta 1}$...O Asn84	94	0
Arg87 $N^{\eta 1}$...O Asp86	100	0
Arg87 $N^{\eta 2}$...O Asn84	5	0

Table II-2. (continued)

Hydrogen bonds observed in the X-ray crystal structure	Percentage of time ensemble in which each hydrogen bond exists	
	MD *in vacuo* (500-1000ps)	MD in water (100-150ps)
Module M5:		
between both termini (a part of β-sheet):		
Leu89 N.....O Tyr97	0	0
Ser91 N.....O Leu95	0	* 80
Trp94 N.....O Ser91	0	6
Tyr97 N.....O Leu89	0	* 92
turn:		
Ser91 O^{γ}...$O^{\delta 1}$ Asp93	* 32	0
Leu95 N....O^{γ} Ser91	0	0
Module M6:		
turn:		
Thr99 O^{γ}....O Thr105	0	0
Asp101 N....O^{γ} Thr99	1	0
Tyr103 N....O^{γ} Thr99	0	0
Gln104 N^{ε}....O Asp101	0	16
Thr105 N...$O^{\delta 1}$ Asp101	0	0
Thr105 O^{γ}..$O^{\delta 1}$ Asp101	0	0
others:		
Arg110 $N^{\eta 2}$...O Ile109	100	0

* conserved hydrogen bonds (see text) which are given in Figures II-2 and II-3.

were not always maintained in the simulations. The helix α1 in M1 and a part of the anti-parallel β-sheet in M3 were maintained as described above. Hydrogen bonds Gln31 N-Lys27 O and Leu33 N-Ala30 O of helix α2 in M2 were retained 96% during the time course. Unexpectedly helix α3 in M2 melted. Hydrogen bonds of the parallel β-sheet in M2 and those of some parts of the anti-parallel β-sheet in M4 and M5 were not conserved.

Module boundaries tend to be in the interior of proteins. Therefore, N- and C-terminal residues of the modules tend to be located closely. However, the N-terminus of the first module and the C-terminus of the last module of the protein are not always buried in the interior, because protein termini are often

exposed to solvent. Therefore, each of the termini of the first and the last modules of a protein are positioned at a distance. Hence some modules at the N- or C-termini of proteins are less compact than the others. The C-terminal module M6, which is less compact in the native state than the other modules, folded into a substantially compact form in the simulation (Fig. II-2 f).

Contrary to modules, all pseudo modules did not maintain native conformations even during 50 picoseconds. Native conformation of a pseudo module is of an extended form, as expected from its definition (Fig. II-2 g). Extended conformations of pseudo modules collapsed into compact forms in the simulations (Fig. II-2 h). RMSD of C^{α} atom positions between their X-ray structures and the averaged conformations of the simulated trajectory were in the range of 5.0 to 8.5 Å (Table II-1).

Qualitatively similar results were obtained from both series of molecular dynamics simulations, using different energy parameters, AMBER and DREIDING. Thus, the results are not affected by particular energy parameters and/or initial conditions.

II-3.2. In water

As in the *in vacuo* simulation, modules M1 to M5 maintained native-like conformations through solution simulations (Fig. II-3 a-e), while all pseudo modules and M6 (Fig. II-3 f) were deformed. Time courses of RMSD of the simulated structures of the modules and pseudo modules from their initial structures (their corresponding X-ray crystal structures) are shown in Figure II-4. The RMSD values for modules M1 to M5 seem to become equilibrated in about 50 picoseconds with asymptotic values of around 2.0 Å. For M3, the RMSD asymptotically approaches 3.0 Å after about 100 picoseconds. The RMSD values for all pseudo modules and M6 seem to become equilibrated in about 100 picoseconds with an asymptotic value of around 4.0 Å. Figure II-5

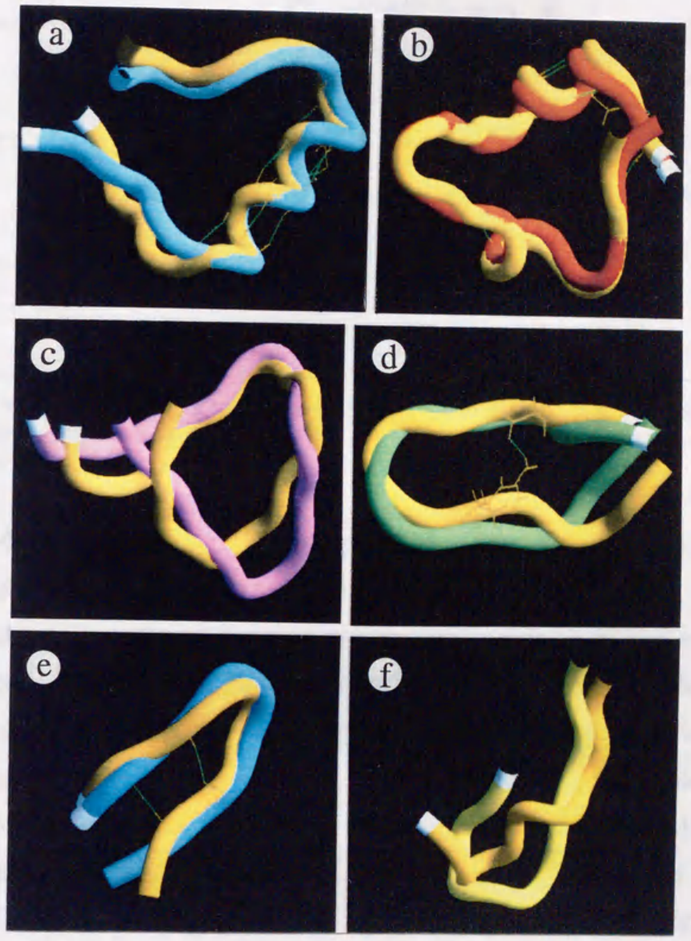

Fig. II-3. The averaged simulation structures of modules in water and their corresponding X-ray crystal structures. Molecular dynamics for 150 picoseconds were performed for each module. Average conformation of each module in the last 50 picosecond trajectories is shown. The presentation by tubes, lines and colouring is the same as for Fig. II-2.

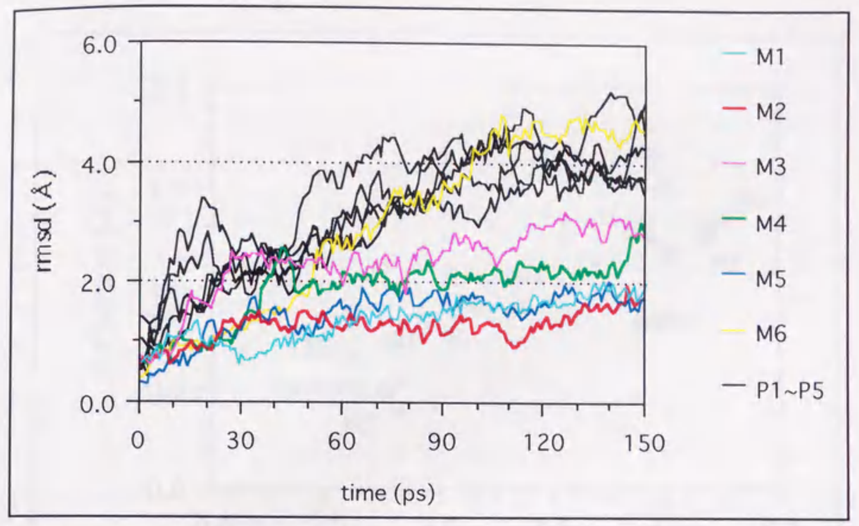

Fig. II-4. Time courses of root mean square deviations (RMSD) of the C^{α} atoms from the corresponding X-ray crystal structures in molecular dynamics simulation in water. The plot is drawn for each module (M1 to M6) and pseudo module (P1 to P5).

shows RMSD values of the averaged structures over the last 50 picosecond trajectories from the X-ray crystal structures in the horizontal axis, and the standard deviations (fluctuations) from the averaged atom positions in the vertical axis. There are essentially two clusters; one consisting of modules M1 to M5 has smaller RMSD (1.2 to 2.7 Å) and fluctuation (0.4 to 0.9 Å), while the other consisting of module M6 and all pseudo modules P1 to P5 has larger RMSD (3.8 to 4.5 Å) and fluctuation (0.8 to 1.5 Å). The RMSD and fluctuation values of the former cluster are comparable with those of barnase (RMSD: 1.5 Å and fluctuation: 0.5 Å). Therefore, the structural stability of M1 to M5 in the 150 picosecond molecular dynamics is comparable with that of barnase.

The conserved hydrogen bonds of the modules during the last 50 picoseconds of the solution simulations are shown in Figure II-3 and Table II-2. Like the simulation *in vacuo*, hydrogen bonds in turns and/or between the N-

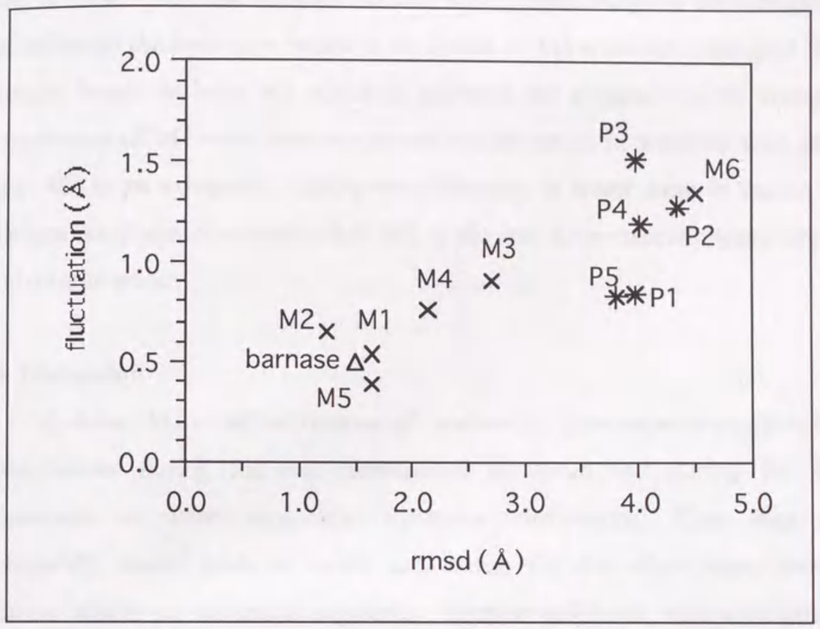

Fig. II-5. Conformational fluctuation vs. RMSD from the X-ray structures for modules (M1 to M6), pseudo modules (P1 to P5) and barnase for the molecular dynamics simulations in water. Average and standard deviations of atomic coordinates were calculated over the conformations in the last 50 picosecond trajectory of each simulation. Mean of the standard deviations of the C^{α} atoms around the averaged positions (fluctuation) is plotted for each module(×) and pseudo module (*) against root mean square deviation (RMSD) of the C^{α} atoms of the averaged structures from the corresponding X-ray structures. For reference, plot of fluctuation vs. RMSD for the last 50 picosecond trajectory of 100 picosecond molecular dynamics simulation of barnase (△) is given.

and the C-termini of the modules, which are important for stabilizing compact conformations, modules, were conserved. Hydrogen bonds Ile25 N-Ser50 O connecting the both ends of M2 and Val36 N-Asn41 O^{δ} around the central turn of M2 were retained for 90 and 74%, respectively, of the last 50 picoseconds of the simulation. For M4, a long-range hydrogen bond Arg83 N^{θ}-Asp75 O^{δ} linking the N- and C-terminal halves of module M4 showed 86% retention. Two of four hydrogen bonds in the anti-parallel β-sheet in M5, Ser91 N-Leu95 O and

Tyr97 N-Leu89 O, were kept for 80 and 92%, respectively, of the simulation time, although the hydrogen bonds in the β-turn of M5 were not conserved. The hydrogen bonds in helix α1 which is essential for maintaining the compact conformation of M1 were well conserved in simulation in water as well as *in vacuo*. M3 is an exception, it behaved differently in water from *in vacuo*; no hydrogen bond was observed within M3 in the last 50 picosecond trajectory of simulation in water.

II-4. Discussion

Modules M1 to M5 of barnase all retained in their respective native-like conformations during the one nanosecond *in vacuo* and during the 150 picoseconds in water molecular dynamics simulations. Thus they are mechanically stable both *in vacuo* and water. On the other hand, pseudo modules, which are extended segments, became deformed with simulations. The compactness of modules is thereby reflected in their conformational stability.

Secondary structures are stabilized mechanically by hydrogen bonds and are therefore regarded as structural elements of globular domains. On the other hand, protein modules are defined as compact structural elements of a globular domain. As in the case of barnase, where the boundaries of M1-M2, M3-M4, M4-M5 and M5-M6 are on the β-sheet, many module boundaries are located on α-helices or β-sheets [2-4,6]. For this reason, the question arose as to whether it was proper to consider modules as structural elements of proteins [29]. Our present study shows that modules are indeed mechanically stable structural elements.

The present study and the close correlation of module boundaries with intron positions in many proteins ([1-10] and unpublished data) suggest that modules behaved as building blocks to produce new proteins through exon

shuffling or fusion, during evolutionary processes [4,12,13]. However, there is no evident correlation between the secondary structure and intron positions [15,16]. As a recruited block, it is essential for each module to be structurally stable so as to maintain a compact conformation in various combinations of modules. The mechanical stability of modules shown in our study supports our hypothesis of protein evolution, from structural and mechanical points of view.

In barnase, hydrogen bonds are localized mainly within modules [22]. In the present simulation about half the number of these hydrogen bonds were conserved (Figs. II-2 a-f, II-3). Hydrogen bonds in turns and/or between the N- and C-termini of each module were conserved.

To determine the main contribution to mechanical stability of modules, we carried out another series of one nanosecond molecular dynamics simulations of barnase modules *in vacuo* with modified potential functions in which ionic charge and hydrogen bond were eliminated [30]. Four of the six modules, M1, M2, M4 and M5, retained their native-like conformations, whereas M3 and M6 became deformed during the simulation. The observed mechanical stability of modules in the absence of ionic charge and hydrogen bond interactions demonstrates that *van der Waals* and non-ionic partial charge interactions are sufficiently strong for most modules to be mechanically stable. Because these interactions are weak per atomic pair and only effective in a short distance, a sufficient number of atomic pairs must be in contact to stabilize the conformations, by only these short-range interactions. In fact the number of atomic pairs interacting in a module is not small because of its compact conformation. Thus, the compactness of modules at the atomic level is essential for mechanical stability of modules.

The isolated modules of barnase maintained their native-like conformations in water for 150 picoseconds, which means that these modules are also mechanically stable in water, but it does not mean that these modules

should form thermodynamically stable native-like conformations in solution, the simulation time is too short to realize a state of thermodynamic equilibrium. Solution structures of isolated M2 and M3 of barnase were studied using 2D-NMR, and some native secondary structures were observed [27]. However, they did not maintain native like 3D structures in solution rather they seemed to have a mixture of conformations. Based on this observation, sufficiently long simulations should lead to large structural deviation of the modules. There would not be any clear difference in thermodynamic stability between modules and pseudo modules. However, we found that there were differences in mechanical stability between modules and pseudo modules. In other words each native conformation of modules was at so deep a local energy minimum that the conformation was retained for at least 150 picoseconds, even if the segment was dissected and isolated in water. On the contrary, each native conformation of pseudo modules was at a shallow local minimum or was not at a local energy minimum so that the conformation was not retained in the 150 picosecond simulation.

The isolated modules in solution are not thermodynamically stable probably because their hydrophobic patch on the module surface involved in module-module interactions in intact barnase becomes exposed to solvent. Contributions of hydrophobic interactions for thermodynamic stability of native conformations of barnase were reported [31]. Hydrophobic cores are mainly formed between modules in the native conformation of barnase [22]. Protein structures and functions are the products of molecular evolution during the past 3.5 billion years. In early periods of evolution, independent primordial modules encoded by mini genes could exist in mechanically and thermodynamically stable conformations in solution. After being recruited into a globular domain, hydrophobic interactions between modules should have been obtained by amino acid replacements to stabilize the assembly. Thus, the thermodynamic

conformational stability of modules isolated in solution would have been lost in the evolutionary process. The compactness and mechanical stability of the original building blocks may well be the remnant still existing as modules in contemporary proteins.

To determine if mechanical stability of modules is common to other proteins, we performed a similar molecular dynamics simulation *in vacuo* for four proteins, myohemerythrin, immunoglobulin heavy chain variable domain, flavodoxin and lysozyme. We selected each of them as representative of α–, β–, α/β– and α+β–type proteins, respectively. As most of the modules were mechanically stable (unpublished data), mechanical stability of modules seems common to a variety of proteins. When designing new proteins, a combination of modules may prove to be a useful technique. The mechanical stability of modules provides insight into such building blocks.

We conclude, from the molecular dynamics simulation, that most modules of barnase are mechanically stable. This was unexpected since these structural units are characterized independently of the secondary structures. The mechanical stability of modules means that a module is a stable structural element of protein architecture. The present study provides a physico-chemical basis for the view that modules were original building blocks of evolving proteins, as well as the correlation between modules and exons observed in various proteins.

II-5. References

1. Gilbert, W.,Glynias, M. On the ancient nature of introns. Gene 135:137-144, 1993.

2. Gō, M. Correlation of DNA exonic regions with protein structural units in haemoglobin. Nature 291:90-92, 1981.

3. Gō, M. Modular structural units, exons, and function in chicken lysozyme. Proc. Natl. Acad. Sci. U.S.A. 80:1964-1968, 1983.

4. Gō, M. Protein structures and split genes. Adv. Biophys. 19:91-131, 1985.

5. Gilbert, W., Marchionni, M.,McKnight, G. On the antiquity of introns. Cell 46:151-153, 1986.

6. Gō, M., Nosaka, M. Protein architecture and the origin of introns. Cold Spring Harb. Symp. Quant. Biol. 52:915-924, 1987.

7. Isaacs, N.W., Machin, K.J.,Masakuni, M. Three-dimensional structure of goose-type lysozyme from the egg white of the Australian black swan, *Cygnus atratus*. Aust. J. Biol. Sci. 38:13-22, 1985.

8. Tittiger, C., Whyard, S.,Walker, V.K. A novel intron site in the triosephosphate isomerase gene from the mosquito *Culex tarsalis*. Nature 361:470-472, 1993.

9. Blake, C.C.F. Exons and the evolution of proteins. Int. Rev. Cytol. 93:149-185, 1985.

10. Gō, M. and Noguti, T. Putative origin of introns deduced from protein anatomy. In: "Tracing Biological Evolution in Protein and Gene Structures." Gō, M., Schimmel, P. (eds.). Amsterdam: Elsevier, 1995:229-235.

11. Patthy, L. "Protein evolution by exon-shuffling." Heidelberg: Springer-Verlag, 1995.

12. Gilbert, W. Why genes in pieces? Nature 271:501, 1978.

13. Blake, C.C.F. Do genes-in-pieces imply proteins-in-pieces? Nature 273:267, 1978.

14. Gilbert, W. The exon theory of genes. Cold Spring Harb. Symp. Quant. Biol. 52:901-905, 1987.

15. Stoltzfus, A., Spencer, D.F., Zuker, M., Logsdon Jr., J.M., Doolittle, W.F. Testing the exon theory of genes: the evidence from protein structure. Science 265:202-207, 1994.

16. Weber, K., Kabsch, W. Intron positions in actin genes seem unrelated to the secondary structure of protein. EMBO J. 13:1280-1286, 1994.

17. Nishimura, S., Nomura, M. Ribonuclease of *Bacillus subtilis*. J. Biochem. 46:161-167, 1959.

18. Mauguen, Y., Hartley, R.W., Dodson, E.J., Dodson, G.G., Bricogne, G., Chothia, C., Jack, A. Molecular structure of a new family of ribonucleases. Nature 297:162-164, 1982.

19. Baudet, S., Janin, J. Crystal structure of a barnase-d(GpC) complex at 1.9 Å resolution. J. Mol. Biol. 219:123-132, 1991.

20. Guillet, V., Lapthorn, A., Hartley, R.W., Mauguen, Y. Recognition between a bacterial ribonuclease, barnase, and its natural inhibitor, barstar. Structure 1:165-177, 1993.

21. Bycroft, M., Ludvigsen, S., Fersht, A.R., Poulsen, F.M. Determination of the three-dimensional solution structure of barnase using nuclear magnetic resonance spectroscopy. Biochemistry 30:8697-8701, 1991.

22. Noguti, T., Sakakibara, H., Gō, M. Localization of hydrogen-bonds within modules in barnase. Proteins 16:357-363, 1993.

23. Yanagawa, H., Yoshida, K., Torigoe, C., Park, J., Sato, K., Shirai, T., Gō, M. Protein anatomy: functional roles of barnase module. J. Biol. Chem. 268:5861-5865, 1993.

24. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., Haak, J.R. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81:3684-3690, 1984.

25. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S.Jr, Weiner, P. A new force field of molecular mechanical simulation of nucleic acids and proteins. J. Am. Chem. Soc. 106:765-784, 1984.

26. Mayo, S.L., Olafson, B.D., Goddard, W.A.III. Dreiding: a generic force field for molecular simulations. J. Phys. Chem. 94:8897-8909, 1990.

27. Ikura, T., Gō, N., Kohda, D., Inagaki, F., Yanagawa, H., Kawabata, M., Kawabata, S., Iwanaga, S., Noguti, T., Gō, M. Secondary structural features of modules M2 and M3 of barnase in solution by NMR experiment and distance geometry calculation. Proteins 16:341-356, 1993.

28. Weiner, S.J., Kollman, P.A., Nguyen, D.T., Case, D.A. An all atom force field for simulations of proteins and nucleic acids. J. Comput. Chem. 7:230-252, 1986.

29. Rogers, J.H. The role of introns in evolution. FEBS Lett. 268:339-343, 1990.

30. Takahashi, K., Gō, M., Noguti, T. Mechanical stability of protein modules determined by molecular dynamics simulations. In: "Tracing Biological Evolution in Protein and Gene Structures." Amsterdam: Elsevier, 1995:175-185.

31. 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, 1992.

32. Kraulis, P.J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Applied Crystallography 24:946-950, 1991.

33. Kabsch, W., Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577-2637, 1983.

CHAPTER III
Essential Factor in Mechanical Stability of Modules

Abstract: Exon-shuffling would yield a stable globular protein, if exons encode stable structural units. Module, a compact segment within a globular protein domain, is considered to be such a structural unit once encoded by an exon, since close correlation is observed between module and exon boundaries in various proteins. In order to elucidate the mechanism of protein evolution and the principle of protein architecture, we investigated the mechanical stability of excised modules. We carried out one nanosecond molecular dynamics simulations of 6 modules of barnase, a bacterial RNase from *Bacillus amyloliquefaciens*, starting with native conformations. We found that all but the C-terminal module, retained native-like conformations through one nanosecond simulations, hence, the modules are mechanically stable. Molecular dynamics simulations with modified nonbonded interactions showed that the main contribution to the stability is due to van der Waals and non-ionic partial charge interactions, indicating that compact conformation of the modules is essential for mechanical stability. Our observations support the view that modules are suitable evolutionary building blocks of proteins.

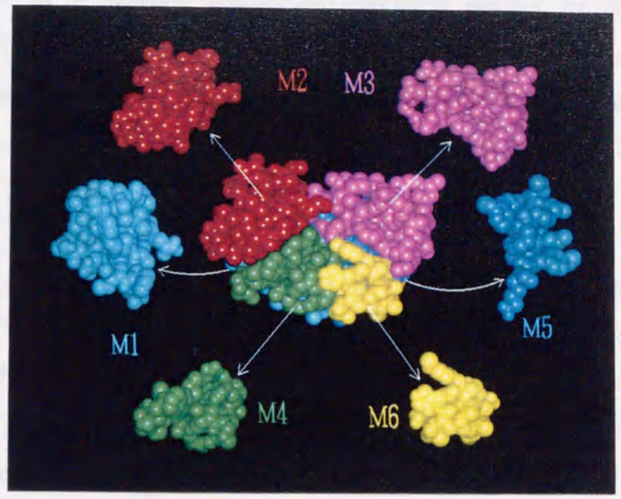

Fig. III-1. Module structure of barnase. Barnase consists of a globular domain, which is decomposed into 6 modules, M1 (cyan), M2 (red), M3 (magenta), M4 (green), M5 (blue) and M6 (yellow).

III-1. Introduction

A globular domain of a protein is decomposed into compact structural units, called modules (Fig. III-1), which are contiguous segments as long as 10 to 40 amino acid residues [1, 2]. Boundaries of modules are determined by the centripetal profile and extension profile [3] based on three dimensional structures of proteins. In various proteins, module boundaries correspond to positions of introns in genes encoding the proteins [1-4]. This suggests that each module was encoded by an exon in early evolutionary periods and that exon shuffling or fusion [4-6] produced various combinations of modules, which resulted in formation of proteins with various functions. The module organization observed in contemporary protein structures implies that each module has retained its primordial conformation through evolutionary processes of various combinations with other modules. If this is indeed the case,

conformation of a module should be specified mainly by intra-module interactions (interactions between atoms within the module) rather than inter-module interactions (interactions between atoms in the module and those in other modules). This point cannot be given too much attention if one is to understand the evolution of protein as well as the principle of protein architecture.

The following suggests the autonomous nature of module conformations: localization of hydrogen bonds within rather than between modules [7], formation of native-like secondary structures in solution structures of isolated modules [8], self-assembly of modules in which each module takes on a stable native-like compact shape [9].

To examine whether native conformations of modules in an existing protein are specified by intra-module interactions, we investigated the mechanical balance in intra-module interactions and the conformational stability of modules by molecular dynamics simulations. Conformation of each module is mechanically stabilized in the native protein, where all forces working on its atoms, including intra- and inter-module interactions and interactions with water molecules, are balanced. If forces from neighboring modules and water molecules are eliminated and only forces of intra-module interactions are left, what will happen to module conformation? This hypothetical state of the module isolated in vacuum can be simulated by molecular dynamics starting from its native conformation (Fig. III-2). The isolated module will retain its native conformation in simulation if forces of intra-module interactions are balanced. Many mechanical properties of a protein molecule, such as conformational changes caused by ligand binding or by amino acid substitution, are characterized by dynamics in the time range of pico to nano seconds [10]. If conformation of the module remains similar to the native conformation for one nanosecond in molecular dynamics simulation, then the native-like

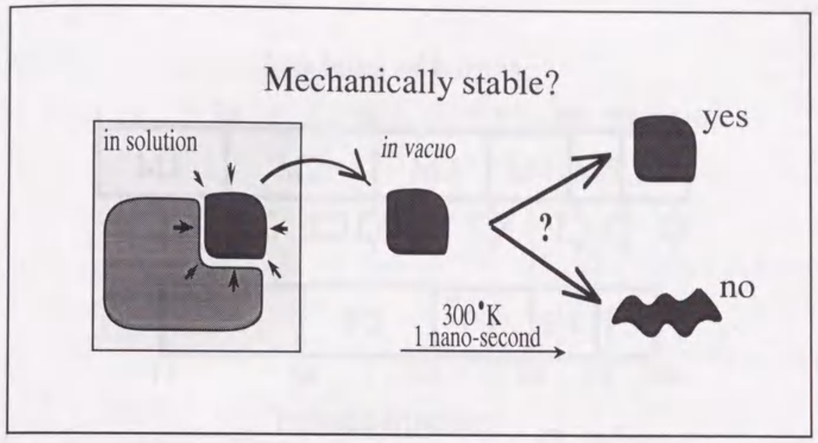

Fig. III-2. Diagram illustrating the simulation study for mechanical stability of modules. In native state of a protein in solution (left on the figure), a module (darker shaded square) interacts with the other part of the protein and solvent. These interactions (represented by short arrows) are eliminated by isolating the module in vacuum (middle). If the native conformation of the module is mechanically stable, the isolated module retains the structure. If not, it is deformed (right). Molecular dynamics simulation of the isolated module elucidates the problem.

conformation of the module is mechanically stable.

Barnase is a bacterial RNase from *Bacillus amyloliquefaciens* [11]. Its tertiary structure has been elucidated by X-ray crystallography [12,13] and NMR spectroscopy [14]. Based on the tertiary structure, barnase was decomposed into six modules M1 to M6 [7] (Fig. III-1). We carried out molecular dynamics simulations of the modules of barnase for a nanosecond and studied their mechanical stability. "Pseudo modules" of barnase were defined as segments starting at a center of one module and ending at the center of the following one (Fig. III-3), and their mechanical stability was also studied by molecular dynamics. In contrast to compact conformations of modules, those of pseudo modules are extended forms in the native protein. We found that conformations of modules were mechanically stable but those of pseudo

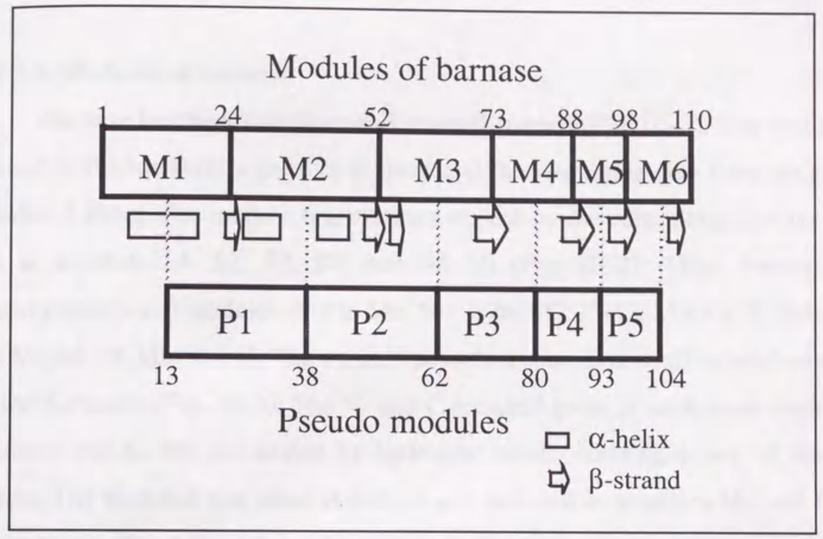

Fig. III-3. Schematic diagram of modules, pseudo modules, and secondary structures of barnase. M1 to M6 and P1 to P5 are modules and pseudo modules, respectively. Residue numbers of boundaries of modules and pseudo modules are shown.

modules were unstable (Takahashi K. *et al.*, in press; see also chapter II). We further examined the origin of the mechanical stability of modules by molecular dynamics simulations. Of the intra-module interactions, nonbonded interactions consisting of electrostatic interactions, hydrogen bonds and van der Waals interactions stabilize module conformations. To investigate the role of these nonbonded interactions in the mechanical stability of modules, we simulated the isolated state of the modules by molecular dynamics calculation with hypothetically modified potential functions where ionic charges and interactions of hydrogen bonds were eliminated.

III-2. Materials and methods

III-2.1. Modules of barnase

Barnase has three α-helices and seven β-strands (Fig. III-3). The first and second β-strands form a parallel β-sheet and the remaining five form an anti-parallel β-sheet. The module boundaries assigned by the centripetal profiles [3] are at residues 24, 52, 73, 88, and 98 [7] (Fig. III-3). Thus, barnase is decomposed into 6 modules, M1 to M6, M1:1-24, M2:25-52, M3:53-73, M4:74-88, M5:89-98, M6:99-110. The module boundaries are almost all located on one of the β-strands (Fig. III-3). The N- and C-terminal parts of each main chain of modules M2 to M5 are linked by hydrogen bonds, forming a part of the β-sheets. The first and two other α-helices are included in modules M1 and M2, respectively. We defined "pseudo modules" of barnase as segments starting at a center of one modules and ending at the center of the following one. There are five pseudo modules, P1 to P5; P1:13-38, P2:39-62, P3:63-80, P4:81-93 and P5:94-104 (Fig. III-3). Each pseudo module includes one or two β-strands.

III-2.2. Simulation methods and energy parameters

Conformations of modules and pseudo modules in the X-ray structure of barnase were energy-minimized, independently. Each polypeptide chain of modules and pseudo modules was terminated at the peptide bond without charged or capped ends. Minimization was performed to find a conformation where the root mean square of atomic forces was less than 0.01 kcal/mol/Å. The energy-minimized conformations were taken for the initial conformations of the molecular dynamics (MD) simulations. The MD simulations were performed on modules and pseudo modules for one nanosecond and 50 picoseconds, respectively. Time steps were 1 femtosecond unless mentioned otherwise. The temperature was maintained at 300 K by coupling (τ = 0.2 picosecond) to an

external bath [15]. Structures were stored every 1 picosecond during the simulations for analysis. In all of minimizations and MD simulations, a linear distance dependent dielectric function, ε=r, was applied and nonbonded interactions were all calculated, without truncation.

To study mechanical stability of each module or pseudo module of barnase, we carried out two MD simulations, which we call "standard MD simulations", using the program and energy parameters of AMBER 3A [16] and those of BIOGRAF [17], respectively. The united-atom model was used in both simulations. Side chains of Lys, Arg, Asp and Glu were ionized. Time step of 0.25 femtosecond was used. To examine the origin of conformational stability of the modules, we carried out the following two MD simulations for each of modules M1 to M5.

1. Charge-reduced MD simulation. In this simulation we used 'the reduced-charge all-atom parameter set' of AMBER 4, in which the net charge of each Lys, Arg, Asp and Glu was reduced to 0.2, and the program of AMBER 3A. We carried out this simulation to investigate effects of ionic charges on mechanical stability.
2. Chargeless, HB-less MD simulation. In this simulation we turned off interactions of hydrogen bonds and set net charge of each Lys, Arg, Asp and Glu to be zero. The energy parameter and program of BIOGRAF for united-atom model were used. Effects of hydrogen bond interactions on the mechanical stability were studied using this simulation.

III-2.3. Conformational comparison

Simulated conformations were averaged over the last half trajectory of each MD simulation. We will hereafter call the averaged conformations "MD-averaged conformations". MD-averaged conformations and the corresponding X-ray structures of modules or pseudo modules were superimposed and

compared. Besides the root mean square deviation of the C^{α}-atom positions (RMSD) between the superimposed structures, we used the root mean square deviation of the C^{α}-C^{α} distances (RMSΔD), defined in the following manner in order to measure conformational differences. The RMSΔD value between an MD- averaged conformation of a module and the corresponding X-ray structure is defined as

$$\mathrm{RMS}\Delta\mathrm{D} = [\, \Sigma\, \Sigma_{\,i-j>4} \{d_M(i,j) - d_X(i,j)\}^2 / n \,]^{1/2},$$

where $d_M(i,j)$ and $d_X(i,j)$ are distances between C^{α} atom of i-th residue and that of j-th residue in the MD-averaged conformation and in the X-ray structure, respectively. The summation is taken over all of the pairs of residues that are separated by more than four residues ($i - j > 4$) and n is the number of such residue pairs. RMSΔD was more suitable than RMSD to measure differences in global main-chain folding patterns of modules.

III-3. Results

III-3.1. Molecular dynamics simulation without modification of nonbonded interactions (standard MD simulation)

Five of the six modules of barnase, M1 to M5, retained native-like conformations through the one nanosecond molecular dynamics simulations (Takahashi K. *et al.*, in press; see also chapter II). The compact foldings of main chains of the modules remained similar to those in the native conformation of barnase through the simulation. The RMSΔD values between the X-ray structures of the five modules and the corresponding MD-averaged conformations were in the range of 2.0 to 3.6 Å (Table III-1). The C-terminal

Table III-1. The root mean square deviation of the C^{α}-C^{α} distances (RMSΔD) in the MD-averaged conformation of each module or pseudo module of barnase in the standard MD simulation from those in the corresponding X-ray structure.

Module	RMSΔD (Å)	Pseudo module	RMSΔD (Å)
M1	2.0	P1	9.1
M2	3.1	P2	10.7
M3	3.6	P3	13.6
M4	2.6	P4	13.4
M5	2.9	P5	8.8
M6	6.9		

module, M6, of which the native conformation is less compact than those of the other modules, folded into a substantially compact form. The RMSΔD value between the X-ray structure and the MD-averaged conformation of M6 was 6.9 Å. All pseudo modules could not maintain their native conformation even for 50 picoseconds, their extended conformations collapsed into compact forms. The RMSΔD values between the X-ray structures and the MD-averaged conformations were in the range of 8.8 to 13.6 Å (Table III-1). These results mean that all modules, except for the C-terminal module M6, are mechanically stable while all pseudo modules are mechanically unstable. The RMSD values of C^{α} atoms between the X-ray structures and the MD-averaged conformations were in the range of 2.3 to 4.4 Å for the five modules M1 to M5, 4.7 Å for M6 and in the range of 5.0 to 8.9 Å for the pseudo modules. Results of the standard MD simulations by the programs and energy parameters of BIOGRAF were essentially the same.

III-3.2. Molecular dynamics simulation with modified nonbonded interactions (charge-reduced MD simulation and chargeless, HB-less MD simulation)

All modules, except for M6, retained their native-like conformations

through the standard MD simulations. Side chains of Asp, Glu, Arg and Lys residues were treated as being fully ionized in the standard MD simulations. Numbers of the charged groups in the modules are given in Table III-2. Effects of the charged groups on the conformational stability of the modules were eliminated by reducing their charges in the charge-reduced MD simulations. The five modules, M1 to M5, retained their native-like conformations throughout a nanosecond in the charge-reduced MD simulations (Fig. III-4). Their RMSΔD values between the MD-averaged conformations of the modules and the corresponding X-ray structures were in the range of 1.3 to 2.5 Å (Table III-2). The RMSD values were 1.5 to 2.8 Å. The secondary structures in the X-ray structures were all observed at the corresponding locations in the MD-averaged conformations of the modules (Fig. III-4). The hydrogen bonds of the α-helix in M1, the two α-helices and the parallel β-sheet in M2, and the parts of the anti-parallel β-sheet formed in M3, M4 and M5 were maintained in the simulated conformations of individual modules. These results of the charge-reduced MD simulations show that the modules are mechanically stable even if charges of ionized groups are reduced.

Hydrogen bonds play essential roles in secondary structures and tertiary structures of proteins. Their contribution to the mechanical stability of module conformations were investigated by the chargeless, HB-less MD simulations in

Table III-2. The root mean square deviation of the C^{α}-C^{α} distances (RMSΔD) in the MD-averaged conformation of each module in the charge-reduced MD simulation and chargeless, HB-less MD simulation from those in the corresponding X-ray structure.

Module	No. of charged groups		RMSΔD	
	+	-	charge-reduced MD (Å)	chargeless, HB-less MD (Å)
M1	1	3	2.1	3.8
M2	3	2	1.7	3.0
M3	5	3	2.5	4.9
M4	2	2	2.2	2.7
M5	1	1	1.3	2.1

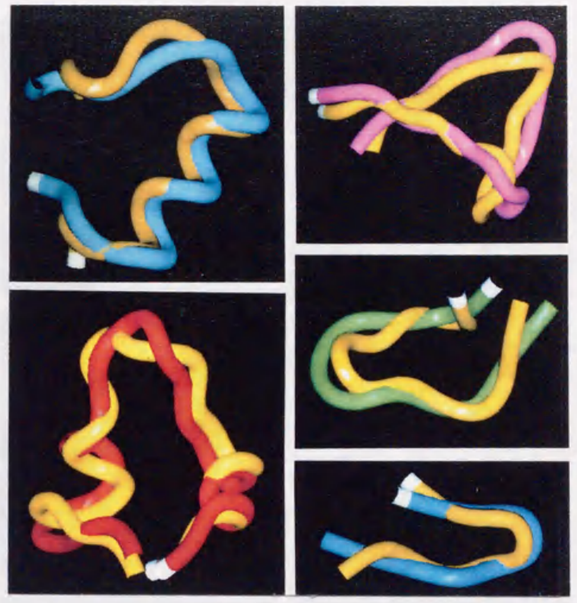

Fig. III-4. Conformations of modules simulated by the molecular dynamics with reduced ionic charges (charge-reduced MD) and their corresponding X-ray structures. The trace of the main chain of each module is shown in a tube model. X-ray structures of modules are colored; M1: cyan, M2: red, M3: magenta, M4: green, and M5: blue. Conformations in the last 500 picosecond trajectory of each molecular dynamics simulation are averaged. The resultant conformation of each module is colored in ocher and superimposed on the corresponding X-ray structure. The N-terminal of each module is indicated in white.

which both the hydrogen bond interactions and electrostatic interactions of charged groups were hypothetically turned off. The result was surprising; four of the five examined modules, M1, M2, M4, and M5 retained their characteristic foldings of the native conformations throughout a nanosecond (Fig. III-5). Module M3 was largely deformed. The RMSΔD values between the MD-averaged conformations of the four modules and the corresponding X-ray structures were 2.1 to 3.8 Å and that of M3 was 4.9 Å (Table III-2). The corresponding RMSD values of the four modules were in 2.4-5.1 Å and that of

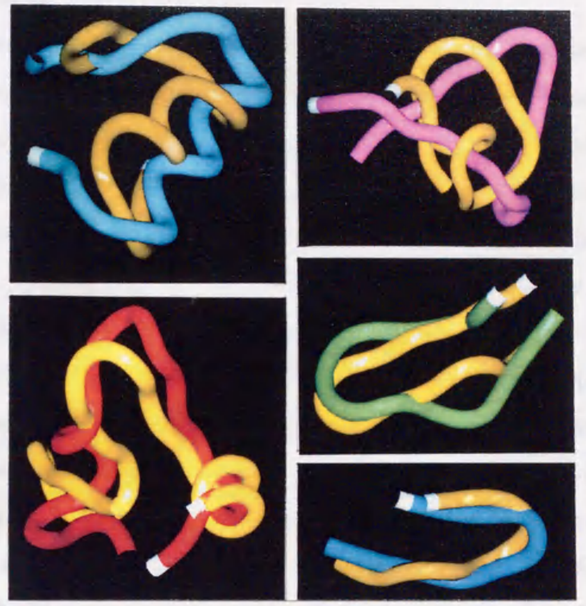

Fig. III-5. Conformations of modules simulated by the molecular dynamics, without ionic charges and hydrogen bond interactions (chargeless, HB-less MD) and their corresponding X-ray structures. The representation is the same as in Fig. III-4.

M3 was 5.0 Å. Some secondary structural features remained in the simulated conformations of the modules despite the absence of hydrogen bond interactions; especially the turns or turn-like structures which are critical to determine the compact native conformations of the modules were retained in M2, M4 and M5. The MD-averaged conformation of M5 is similar to the native conformation, an anti-parallel β-sheet. Modules M1 and M2 formed native-like helical structures, however, their length and location were not identical to the native ones. These results of chargeless, HB-less MD simulations show that most of the modules are mechanically stable even if hydrogen bond interactions are eliminated.

III-4. Discussion

We simulated the hypothetical state of barnase modules isolated in vacuum by one nanosecond molecular dynamics. All but the C-terminal module retained native-like conformations through the one nanosecond simulations. This means that atomic interactions are balanced within each of the modules and thus conformations of the modules are mechanically stable (Takahashi K. *et al.*, in press; see also chapter II) and it also suggests that tertiary structures of modules are determined mainly by intra-module interactions.

We further examined roles of nonbonded interactions in the conformational stability of the modules by simulating their isolated state with modified potential functions where ionic charges and hydrogen bond interactions are eliminated. We found that conformations of the modules remain stable without the hydrogen bond interactions and electrostatic interactions of ionic charges. Therefore, other nonbonded interactions, the van der Waals interaction and the electrostatic interaction between partial charges of non-ionic groups, mainly contribute to the mechanical stability. These nonbonded interactions are weak for each atomic pair and are only effective in case of a short distance. A sufficient number of atoms should be in contact to stabilize the conformations. In general, the more compact conformation has a larger number of atomic contacts. Modules are so compact as to have a large number of atomic contacts. Therefore, it is concluded that compact conformations of modules are essential for mechanical stability. In contrast to modules, pseudo modules are not compact and do not have a large number of atomic contacts, thus their conformations are mechanically unstable. Hydrogen bonds, however, play important role in the mechanical stability of modules. Secondary structures are retained better in charge-reduced MD simulation (Fig. III-4) than in chargeless, HB-less MD simulations (Fig. III-5), where interactions of hydrogen bonds are

turned off, hydrogen bond interactions contribute to stabilization of local secondary structures.

To examine the mechanical balance in intra-module interactions, we simulated a hypothetical state of an isolated module in vacuum. In an actual situation, modules are integrated into a protein surrounded by water molecules; surface of modules is considerably exposed to water. Electrostatic interactions and hydrogen bonds between exposed atoms to water are weakened by hydration effects. In vacuum simulations, effects of electrostatic interactions and hydrogen bonds between atoms on module surface might be more potent than in the actual situation. However, this does not seem to affect the results of our simulations. We found that most of the modules investigated are mechanically stable in the complete absence of ionic charges and hydrogen bonds (Fig. III-5).

We carried out the same simulation of modules of four other proteins: myohemerythrin, immunoglobulin V_H domain, flavodoxin and lysozyme, which are α-, β-, α/β- and $\alpha+\beta$-type proteins, respectively. Most of the modules were mechanically stable.

Correlation between intron positions and module boundaries [1-4] suggests that there was a one-to-one correspondence between exons and modules in the early evolutionary period and that exon shuffling or fusion [4-6] produced various combination of modules, which resulted into the creation of stable proteins with various functions. It is essential that each module be structurally stable and maintain a compact conformation in various combinations of modules. We showed that most of the modules are mechanically stable. This result further supports the view that modules are indeed evolutionary building blocks of proteins. Assembly of mechanically stable modules are likely to result in mechanically stable proteins. Modification of a protein by exchanging modules is also likely to be successful. Alternative

splicing produces several combinations of exons from a single copy gene, depending on cell types or stages of development. The resultant proteins often have slightly tuned functions. The mechanical stability of modules is thought to be important in this modification of protein functions.

III-5. References

1. Gō, M. Correlation of DNA exonic regions with protein structural units in haemoglobin. Nature 291:90-92, 1981.

2. Gō, M. Modular structural units, exons, and function in chicken lysozyme. Proc. Natl. Acad. Sci. U.S.A. 80:1964-1968, 1983.

3. Gō, M., Nosaka, M. Protein architecture and the origin of introns. Cold Spring Harb. Symp. Quant. Biol. 52:915-924, 1987.

4. Gō, M. Protein structures and split genes. Adv. Biophys. 19:91-131, 1985.

5. Gilbert, W. Why genes in pieces? Nature 271:501, 1978.

6. Blake, C.C.F. Do genes-in-pieces imply proteins-in-pieces? Nature 273:267, 1978.

7. Noguti, T., Sakakibara, H., Gō, M. Localization of hydrogen-bonds within modules in barnase. Proteins 16:357-363, 1993.

8. Ikura, T., Gō, N., Kohda, D., Inagaki, F., Yanagawa, H., Kawabata, M., Kawabata, S., Iwanaga, S., Noguti, T., Gō, M. Secondary structural features of modules M2 and M3 of barnase in solution by NMR experiment and distance geometry calculation. Proteins 16:341-356, 1993.

9. Yoshida, K., Shibata, T., Masai, J., Sato, K., Noguti, T., Gō, M., Yanagawa, H. Protein anatomy: spontaneous formation of filamentous helical structures from the N-terminal module of barnase. Biochemistry 32:2162-2166, 1993.

10. Karplus, M., Petsko, G.A. Molecular dynamics simulations in biology. Nature 347:631-639, 1990.

11. Nishimura, S., Nomura, M. Ribonuclease of *Bacillus subtilis*. J. Biochem. 46:161-167, 1959.

12. Mauguen, Y., Hartley, R.W., Dodson, E.J., Dodson, G.G., Bricogne, G., Chothia, C., Jack, A. Molecular structure of a new family of ribonucleases. Nature 297:162-164, 1982.

13. Baudet, S., Janin, J. Crystal structure of a barnase-d(GpC) complex at 1.9 Å resolution. J. Mol. Biol. 219:123-132, 1991.

14. Bycroft, M., Ludvigsen, S., Fersht, A.R., Poulsen, F.M. Determination of the three-dimensional solution structure of barnase using nuclear magnetic resonance spectroscopy. Biochemistry 30:8697-8701, 1991.

15. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., Haak, J.R. Molecular dynamics with coupling to an external bath. J . Chem. Phys. 81:3684-3690, 1984.

16. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. Jr, Weiner, P. A new force field of molecular mechanical simulation of nucleic acids and proteins. J. Am. Chem. Soc. 106:765-784, 1984.

17. Mayo, S.L., Olafson, B.D., Goddard, W.A.III. Dreiding: a generic force field for molecular simulations. J. Phys. Chem. 1990;94:8897-8909.

CHAPTER IV

Design of Mini-Barnase by Removing a Module

Abstract: A globular domain, a folding unit of a protein, is decomposed into compact modules which are contiguous segments of 10-40 amino acid residues. Correlation between modules and exons suggests that modules were primordial polypeptides encoded by ancestral exons and that modules were combined together to yield present-day globular domains. The above view of protein evolution gives us an idea of designing a new protein by a new combination of modules. To establish the module-based protein design, we have to elucidate the role of modules in protein architecture. Design of a mini-protein created by removing one module from an existing protein will give us insight into the evolutionary combination process of modules. We have designed a mini-barnase by removing the second module M2 from barnase. Barnase is a single domain bacterial RNase from *Bacillus amyloliquefaciens*. Molecular dynamics study showed the conformation of the molecular-modeled mini-barnase was mechanically stable. Based on a theoretical prediction method, two amino acid residues were replaced in order to increase the thermodynamic stability and solubility of the mini-barnase. The mini-barnase site-specifically ^{15}N-labeled was chemically synthesized and its conformation was examined by NMR

measurement. The result of the NMR experiment indicated the mini-barnase took a specific stable conformation which appeared to be similar to the designed. The solubility was high enough for the 3D structure determination by NMR.

IV-1. Introduction

To make a desired functional protein by designing its structure based on the principle of protein architecture is a goal of protein engineering. In spite of efforts to elucidate the principle in protein design, our knowledge is still limited. Reflecting this situation most of successful approaches in protein design are modification of local structures of existing proteins by replacing one or a few amino acid residues [1-4]. Alternative approach, so called, *de novo* design, has been tried by several groups. In this approach an amino acid sequence of a polypeptide chain is designed so that the polypeptide chain can fold into a desired template structure. Some β-hairpins [9,10] and bundle structures of α-helices [5-8] were successfully designed in this approach. Other polypeptides designed to fold into mixed α/β- or all β-structures were shown to possess expected secondary structural features but to have molten globule-like structures rather than rigid tertiary structures [11-14].

The ancestral proteins of present-day proteins have been made in the course of early biological or pre-biological evolution. Consideration of protein evolution process would give us a valuable hint on protein design.

A globular domain of a protein consists of compact structural units, modules, which are contiguous segments of 10-40 amino acid residues [15,16]. Boundaries of modules have been found to have good correlation with positions of introns in genes of various proteins, including hemoglobin [15,17,18], lysozyme [16,19], triose phosphate isomerase [20-23] and so forth [24]. The correlation suggests that modules were encoded by their corresponding exons in

ancestral genes, and various globular domains evolved by assembling the preexisting modules through exon shuffling [15,16,20,24-28].

Barnase, a bacterial RNase from *Bacillus amyloliquefaciens*, consisting of a single globular domain of 110 amino acid residues [29,30] was decomposed into six modules, M1 to M6 [31]. We have examined atomic interactions in the 3D structure of barnase and have found hydrogen bonds are localized mainly within each of the modules [31]. Solution structures of chemically synthesized modules M2 and M3 have been studied by 2D NMR and secondary structures have been observed in the respective modules at the similar positions to those in barnase [32]. Synthesized polypeptides of module M1 have been found to assemble together in aqueous solution and form a stable filamentous structure. Scanning tunneling microscopy has shown each M1 module in the assembly to take on a compact conformation similar to that in barnase [33]. These facts observed in the studies of the modules of barnase indicate that a compact conformation of a module is specified predominantly by interactions within the module and the conformation is further stabilized by association with other modules. Thus modules could have behaved as building blocks of globular domains in protein evolution.

If the above view of protein evolution is true, it will be possible to make an artificial protein by removing one or more modules from a globular domain of a present-day protein; the polypeptide chain consisting of the remaining modules will be able to fold into a stable conformation similar to that of the intact protein except for the removed parts. Design of such a mini-protein will make the first step in the establishment of a basis of a new method of creating new functional proteins by assembling the evolutionary building blocks, modules.

We designed a mini-protein by removing a module M2 from barnase. The polypeptide chain of the mini-protein, named mini-barnase, was chemically

synthesized and its solution structure was examined by 2D NMR. Result of the NMR study indicated that the mini-barnase folded into a specific stable conformation which appeared similar to the native conformation of barnase except for the removed module. In this paper we describe how the mini-barnase was designed theoretically and we report the results of the experimental examination of the designed mini-barnase.

IV-2. Materials and methods

IV-2.1 Modules of barnase

Conformation of barnase has been elucidated by X-ray crystallography [34-36] and NMR spectroscopy [37]. Barnase has three α-helices, a two-stranded parallel β-sheet and a five-stranded anti-parallel β-sheet. Boundaries of the modules assigned by the centripetal profile [20] are at residues 24, 52, 73, 88 and 98 [31]. Thus the 6 modules are M1:1-24, M2:25-52, M3:53-73, M4:74-88, M5:89-98 and M6:99-110(Fig. IV-1 a). Modules M1 and M2 have one and two α-helices, respectively. Module M2 forms the parallel β-sheet near its N- and C-terminal ends.

IV-2.2 Design of mini-barnase

Molecular modeling of mini-barnase

The conformation of the mini-barnase was modeled in the following procedures by using the program BIOGRAF(Version 3.21, Molecular Simulations Inc., Burlington, MA, 1992) [38]. First, the segment 26-51 was removed from the crystal structure of barnase (Brookhaven Protein Data Bank [39], entry 1RNB) [35]. Through this paper the same numberings of residues and modules are used for the mini-barnase as those for barnase to see the

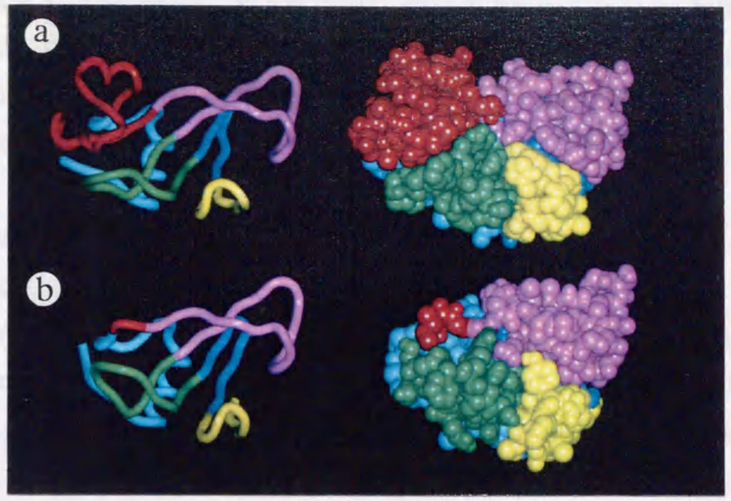

Fig. IV-1. The conformations of barnase and the designed mini-barnase. Modules are shown in different colors: M1 is in sky blue, M2 red, M3 magenta, M4 green, M5 blue and M6 yellow. **a**: Conformation of barnase [35] represented in tube model (left) and space-filling model (right) is shown in the direction of its catalytic sites. **b**: Conformation of the molecular-modeled mini-barnase (unmodified) by removing module M2, sampled at 115 ps in the molecular dynamic simulation. N- and C- terminal residues of M2 left in the mini-barnase are shown in red.

correspondence easily. Second, a peptide bond was made between residues 25 and 52 by the molecular modeling method in which three residues on both the N- and C-terminal sides of the bond were relaxed by energy minimization. Neither irregular geometry of chemical bonds nor collision of atoms were found in the modeled conformation. Third, the conformation of the first residue and that of the second side chain, which were lacked in the crystal structure of barnase used in this study, were modeled by using the atomic coordinates of the crystal structure of the complex of barnase and barstar (Brookhaven Protein Data Bank, entry 1BGS) [36] after superimposing the alpha carbon, carbonyl carbon and amide nitrogen atoms of the second residue Glu 2 in the two crystal

structures of barnase. Ile 25 and Phe 82 of the mini-barnase were replaced by Thr and Ala, respectively. In this paper we call the mini-barnase before the modification by these amino acid replacements "mini-barnase(unmodified)" and the mini-barnase after the modification "mini-barnase(modified)", if we need to discriminate the two versions of the mini-barnase.

Molecular dynamics simulations

Four layers of water molecules were put on the modeled mini-barnase(unmodified) and energy of the water and protein system was minimized until the root-mean-square of atomic forces (RMSF) got less than 0.01 kcal mol^{-1} $Å^{-1}$. A molecular dynamics simulation was performed on this system for 115 ps. The temperature was maintained at 298 K by the heat-bath-coupling method (τ=0.2 ps) [40]. Time step was 1 fs. In minimization and molecular dynamics simulation a linear distance dependent dielectric function, ε=r, was applied and nonbonded interactions were truncated at 9 Å distance. Hydrogen atoms which can form hydrogen bonds were treated explicitly but other hydrogen atoms were united to the heavy atoms. Side chains of Lys, Arg, Asp and Glu, and both ends of the main chain were ionized. During the simulation some of the water molecules vapored out of the water shell. After execution of the first 25 ps simulation, we filled the defects of the four-layered water shell. The water shell was equilibrated by 0.8 ps molecular dynamics simulation after 200 cycles of conjugate gradient minimization, and then the simulation was continued. Supplement of water molecules was repeated in this way every 10 ps after the first 25 ps. In this simulation we used the program and energy parameters of BIOGRAF.

The molecular dynamic simulation of the mini-barnase(modified) was carried out in the same manner as that of the mini-barnase(unmodified) except for the simulation length , 75 ps on this system.

Prediction of contribution of a residue to thermodynamic conformational stability and solubility in aqueous solution

Ooi *et al*. have developed a method of predicting thermodynamics of protein unfolding, whereby Gibbs free energy change in protein unfolding, ΔG^U, is calculated as a linear function of changes of atomic accessible surface areas (ASAs) from those of the folded conformation to those of the extended conformation [41-44]. By using their method, we predicted contributions of residues to ΔG^U for the mini-barnase. Contribution of residue i to ΔG^U was calculated by the following equation

$$\Delta g_i^U = \Sigma_j \ (g_{jh} + g_{jc})(ASA_j^U - ASA_j^F) \qquad (1)$$

, where Δg_i^U is the contribution of residue i, g_{jh} and g_{jc} are the hydration and chain free energy per unit of ASA of atom j, respectively, and ASA_j^U and ASA_j^F are the values of ASA of atom j in unfolded and folded states, respectively. The summation is done over all heavy atoms in residue i. No hydrogen atom was treated explicitly in the calculation of atomic ASAs. ΔG^U is the sum of Δg_i^Us over all residue i's. Ooi *et al*. evaluated hydration and chain free energy per unit of ASA for seven atomic groups, aliphatic carbon, aromatic carbon, hydroxyl oxygen, amide nitrogen, carbonyl carbon, carbonyl oxygen and sulfur, by least-squares fitting to experimental free energy changes of solution of small monofunctional molecules and those of unfolding of proteins, assuming the linear relation between changes of free energies and changes of atomic ASAs [42-44]. The same set of the values at 298 K was used here. Atomic ASAs in the unfolded state were represented by those of the extended polypeptide chain with the amino acid sequence of the mini-barnase as they were in the studies by Ooi and Oobatake [41,42]. The *van der Waals* radii of heavy atoms and that of water

were the same as those in their studies. They used the X-ray structures of proteins for the calculation of the atomic ASAs in the folded states of proteins. Since we were to design a new protein whose X-ray or NMR structure did not exist, we used the conformations in the records of the molecular dynamics simulations of the molecular-modeled mini-barnase. Values of ASA of atom j in conformations sampled every 0.1 ps in a period of 5 ps were averaged to represent $ASAj^F$ in the period of the simulation. The solubility of a protein is determined by Gibbs free energy change in aggregation, ΔG^A. We predicted contribution of residue i to ΔG^A by using the following equation

$$\Delta gi^A = \Sigma_j \ (g_{jh} + g_{jc})(ASAj^A - ASAj^F) \qquad (2)$$

, where Δgi^A is the contribution of residue i, and $ASAj^A$ is the value of ASA of atom j in aggregated state. In aggregated sate, the molecules assemble together and their access to water molecule is interrupted. Therefore, we assumed that atomic ASAs were essentially zero in the aggregated states, *i.e.* $ASAj^A = 0$. ΔG^A is the sum of Δgi^As over all residue i's.

The calculation of atomic ASA by the method of Shrake and Rupley [45] has an error caused by the incompleteness of uniform distribution of 92 points over the surface of a sphere. The error is not negligible in the calculation of the free energies [41]. We increased the accuracy of the calculation of atomic ASA by the following procedure. We generated one hundred different sets of atomic coordinates of a molecule with a fixed conformation by rotating the coordinate system randomly, and calculated atomic ASAs for each set of the coordinates by the method of Shrake and Rupley. Then values of respective atomic ASAs which were precise enough for the calculation of the free energies were obtained by averaging the hundred sets of the calculated ASA values.

Accessibility of a residue is the ratio between the sum of ASAs of atoms

of the residue in the folded conformation and that in the extended form [46]. This value indicates whether the residue is on the surface or in the interior of the molecule.

IV-2.3 Chemical synthesis and NMR spectroscopy of mini-barnase

Boc-amino acid derivatives, reagents for solid-phase peptide synthesis and ^{15}N-labeled amino acids were purchased from Peptide Institute Inc. (Osaka, Japan), Watanabe Chemical Ind. Ltd. (Hiroshima, Japan) and Nihon Shoji Co., Ltd. (Osaka, Japan), respectively.

The chemical synthesis of the designed mini-barnase(modified) was carried out using the previously described procedure [47]. Briefly, partially protected peptide thioesters corresponding to mini-barnase (1-25+52), (53-81) and partially protected peptide (82-110) were prepared via a solid-phase method using Boc-amino acids. For the introduction of the residue Ala11, Tyr13, Leu14, Leu89, Tyr90, Ser91, Ser92, Asp93, Leu95, and Tyr97, ^{15}N-labeled amino acid derivatives were used. The peptide segments were successively condensed by the activation of thioester groups by silver ions to give a protected form of mini-barnase. After deprotection, the mini barnase was highly purified by HPLC. The purity was confirmed by analytical HPLC, amino acid analysis, and matrix-assisted laser desorption ionization (MALDI) mass spectrometry.

Lyophilised mini-barnase was dissolved in 90% H_2O / 10% D_2O. Acids which had remained still after the purification was excluded by ultrafiltration until the pH of the solution rose to 5.1. Simultaneously mini-barnase was concentrated to 0.1 mM. Two dimensional ^{1}H-^{15}N heteronuclear single-quantum coherence (HSQC) spectrum [48] was measured at 25 °C on Varian Unity-plus 750 spectrometer. The resonance of the solvent water was suppressed by presaturation during the relaxation delay. The spectrum consisted of 128 complex t1 points with each free induction decay having 2048 complex

t2 points. 160 transients were acquired per t1 value to gain a signal to noise ratio. The spectral width for ^{1}H observation was 12000 Hz and that for ^{15}N was 3000 Hz. The spectrum was zero-filled to 2048 x 512 real points and was processed with shifted sinebell window function in both dimensions. Linear prediction was applied to raise the spectral resolution of ^{15}N dimension. ^{1}H and ^{15}N chemical shifts are referenced to external TSP (0 ppm) and NH_4Cl (24.93 ppm), respectively.

IV-3. Results

IV-3.1 Design of mini-barnase

We designed a mini-protein by removing module M2 from barnase. Based on the following consideration, we selected the module M2 to be removed. Barnase consists of six modules, M1 to M6. Each module interacts with other modules to stabilize the 3D structure of barnase. Among the six module, M2 forms the least hydrogen bonds and hydrophobic contacts with the remaining part of the protein [31,49]. Therefore, the effects of excision of one module on the conformational stability of the remaining portion will be maintained minimum when M2 is removed from barnase.

In designing the mini-barnase lacking module M2, we considered the following points: 1) conformationally appropriate connection of the chains on either side of module M2 in the molecular modeling of the mini-barnase, 2) mechanical stability of the conformation of the molecular model, and 3) thermodynamic stability of the conformation and solubility of the molecule.

Molecular modeling of mini-barnase

If we cut out the segment 25-52, which corresponds exactly to module M2, from barnase, there remain two chains, M1 and M3-M6 chains. In the

conformation of barnase, the carbonyl carbon atom of the C-terminus of M1 is 6.5 Å away from the amide nitrogen atom of the N-terminus of M3. If the two atoms are connected chemically, a large stress will act on the conformations of the M1 and M3-M6 chains and make them unstable.

In general two ends of a module are relatively close to each other, because module has a compact conformation. We analyzed the barnase 3D structure and examined distances between carbonyl carbon atoms in the vicinity of the boundary of M1 and M2, and amide nitrogen atoms in the vicinity of the boundary of M2 and M3. The distance between the carbonyl carbon atom of Ile 25 and the amide nitrogen atom of Gly 52 was the shortest, 3.9 Å. Removing the segment 26-51 instead of the segment exactly corresponding to M2 and connecting the residues 25 and 52, we could make a molecular model of single polypeptide chain of the mini-barnase lacking one module M2 of barnase (Fig. IV-1 b). Strictly speaking, we left one residue of each end of M2 in the mini-barnase so that the two chains left after the excision of the module should be connected smoothly. Since this discrepancy between module M2 and the actually removed segment is small and the identification of module boundaries itself has error in the range of a few residues, we do not think this discrepancy critical and also call the removed segment module M2 in this paper.

Mechanical stability of conformation of mini-barnase

Conformation of a native protein is mechanically stable, because the balance of all forces working on its atoms is kept in small amplitude thermal perturbations. The absence of the building block consisting of 26 residues might break the balance of atomic forces and make it difficult for the mini-barnase to retain its conformation. We examined the mechanical stability of the modeled mini-barnase conformation by molecular dynamics study. The conformation of the mini-barnase(unmodified) was retained during 115 ps in water environment,

except for a loop of module M4(Fig. IV-1). C^{α} atoms in the conformation on the trajectory of the simulation at 115 ps were superimposed on the corresponding atoms in the 3D structure of barnase, except for those in the loop from 79 th to 83 th residues. The root-mean-square deviations (RMSD) of the superimposed C^{α} atoms was 1.3 Å. Thus, the conformation of modeled mini-barnase is essentially mechanically stable. For C^{α} atoms in the loop the RMSD value was 4.7 Å. The loop moved in the direction to narrow the cleft between module M1 and M4(Fig. IV-1).

Amino acid replacements to increase thermodynamic conformational stability and solubility of the mini-barnase

A reliable method of predicting thermodynamic stability of a protein conformation has not been established yet. Therefore, whether the designed mini-barnase takes an expected conformation or not should be examined by experimental works. What we can do in the process of designing the mini-barnase is to select out amino acid residues that would decrease thermodynamic conformational stability or solubility of the molecule, and to replace them by more suitable ones. High solubility into aqueous solution is required for 3D structure determination by NMR, which we are going to do.

After the excision of the module M2, the interface between M2 and other modules in barnase becomes a part of the molecular surface of mini-barnase. Twelve residues of the mini-barnase(unmodified) increased their accessibilities by more than 0.1 from those of barnase. Out of the twelve residues, four were hydrophobic: Val 3, Ile 25, Tyr 78 and Phe 82(Fig. IV-2). Exposure of a hydrophobic residue on protein surface often decreases thermodynamic conformational stability and/or solubility of the protein in water. We focused our attention on the four hydrophobic residues and predicted their contributions to the conformational stability and the solubility of the mini-barnase.

Considering conformational changes in the vicinities of the four residues during the 115 ps simulation, we calculated two sets of the averaged atomic ASAs in the folded state; one was a set of averages of respective atomic ASAs in conformations in the period of 80-85 ps and the other in the period of 110 -115 ps. Then we calculated two sets of their contributions to Gibbs free energy change in unfolding, Δg_i^U, and those to Gibbs free energy change in aggregation, Δg_i^A(Table IV-1).

Values of $\Delta g_{Phe82}{}^U$, the predicted contribution of Phe 82 to ΔG^U(Gibbs free energy change in unfolding), are both negative, -0.31 and -0.37 kcal mol^{-1}, respectively; the one is calculated by using the averaged atomic ASAs in the period of 80-85 ps and the other in the period of 110-115 ps. This means Phe 82 would destabilize the folded conformation of the mini-barnase. On the other hand values of $\Delta g_{Tyr78}{}^U$ are both positive, 1.13 and 1.11 kcal mol^{-1}, indicating

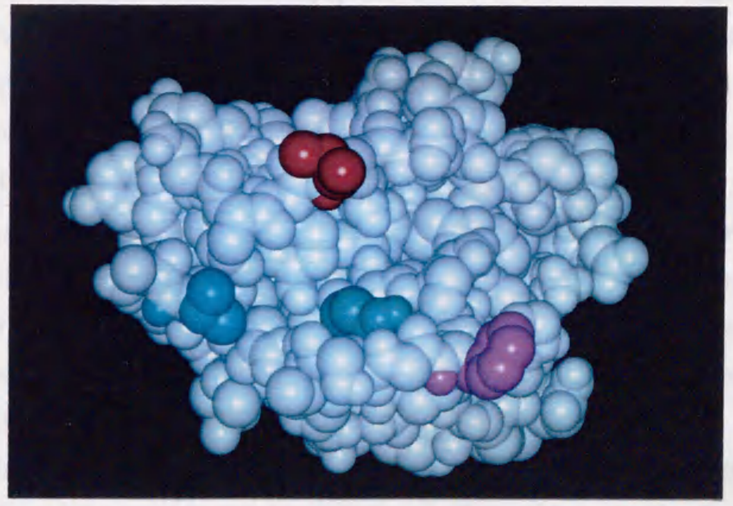

Fig. IV-2. The four hydrophobic residues of which accessibilities increased more than 0.1 after the removal of M2; Val 3 is sky blue, Ile 25 red, Tyr 78 right green and Phe 82 magenta. The mini-barnase is shown in the the direction of the left side of the Fig. IV-1 b.

Table IV-1. Predicted contributions of the four hydrophobic residues to gibbs free energy changes in unfolding and aggregation of the mini-barnase before the amino acid replacement.

Residue *i*	Contribution to ΔG^{U*} Δg_i^U (kcal mol^{-1})		Contribution to ΔG^{A*} Δg_i^A (kcal mol^{-1})	
	80-85ps†	110-115ps†	80-85ps†	110-115ps†
Val 3	-0.03	-0.02	0.17	0.18
Ile 25	-0.23	-0.06	-0.15	0.01
Tyr 78	1.13	1.11	0.00	-0.02
Phe 82	-0.31	-0.37	-3.12	-3.19

*ΔG^U and ΔG^A are Gibbs free energy changes in unfolding and aggregation of the mini-barnase, respectively.

†Values in the columns of 80-85 ps are calculated by equations (1) and (2), where atomic ASAs in the folded state are averages of respective atomic ASAs in conformations in the period of 80-85 ps in the molecular dynamics of the mini-barnase before the amino acid replacement. Values in the columns of 110-115 ps are calculated in the same way by using conformations in the period of 110-115 ps.

Tyr 78 stabilizes the folded conformation. Values of $\Delta g_{Phe82}{}^A$, the predicted contribution of Phe 82 to ΔG^A(Gibbs free energy change in aggregation), are -3.12 and -3.19 kcal mol^{-1}, and imply this residue would facilitate aggregation of the molecules. Tyr 78 is considered to have little effect on the solubility, because $\Delta g_{Tyr78}{}^A$ are small in both the periods. In each of the periods, $\Delta g_{Val3}{}^U$ is negative but small and $\Delta g_{Val3}{}^A$ is slightly positive. Val 3, therefore, causes probably neither conformational instability nor aggregation. Values of $\Delta g_{Ile25}{}^U$ and $\Delta g_{Ile25}{}^A$ are -0.23 and -0.15 kcal mol^{-1}, respectively in the period of 80-85 ps, while they are negligibly small in the period of 110-115 ps. Ile 25 could be a factor to decrease thermodynamic stability and the solubility. From the above consideration, we replaced Ile 25 and Phe 83 by threonine and alanine, respectively.

In the final process of the designing the mini-barnase, molecular dynamics of the mini-barnase(modified), of which two amino acid residues were replaced, was carried out for 75 ps to examine the mechanical stability and to predict the effect of the amino acid replacement on the thermodynamic

properties. The amino acid replacement of Ile 25 to Thr and Phe 82 to Ala did not change the mechanical stability of the conformation of the mini-barnase. The conformation was retained during the 75 ps simulation. The RMSD value of corresponding C^{α} atoms, except for the loop in M4, was 1.3 Å between the simulated mini-barnase conformation at 75 ps and the X-ray structure of barnase. For C^{α} atoms in the loop, the value was 5.1 Å. Contributions of Thr 25 and Ala 82 to ΔG^{U} and ΔG^{A} were calculated from the simulated conformations in the periods of 40-45 ps and 70-75 ps (Table IV-2). Because the predicted values are all positive, the amino acid replacement of Ile 25 to Thr and Phe 82 to Ala was expected to increase the thermodynamic stability and the solubility of the mini-barnase. The mechanical stability of the mini-barnase(modified) has been confirmed further by 100 ps molecular dynamics study by using another set of program and energy parameters of DISCOVER(version 2.95, Biosym Technologies, Inc., San Diego, CA, 1994) [50]. The RMSD values of corresponding C^{α} atoms except for the loop in M4, and C^{α} atoms in the loop were 1.7 Å and 5.5 Å, respectively, between the simulated mini-barnase conformation at 100 ps and the X-ray structure of barnase.

Table IV-2. Predicted contributions of the two hydrophobic residues to gibbs free energy changes in unfolding and aggregation of the mini-barnase after the amino acid replacement.

Residue *i*	Contribution to ΔG^{U*} Δg_i^{U} (kcal mol^{-1})		Contribution to ΔG^{A*} Δg_i^{A} (kcal mol^{-1})	
	40-45ps†	70-75ps†	40-45ps†	70-75ps†
Thr 25	0.12	0.11	0.52	0.52
Ala 82	0.05	0.06	0.18	0.18

*See footnote of Table IV-1.

†Values in the columns of 40-45 ps are calculated by equations (1) and (2), where atomic ASAs in the folded state are averages of respective atomic ASAs in conformations in the period of 40-45 ps in the molecular dynamics of the mini-barnase after the amino acid replacement. Values in the columns of 70-75 ps are calculated in the same way by using conformations in the period of 70-75 ps.

IV-3.2 NMR Measurement of solution structure of mini-barnase

Figure IV-3 shows the ^{1}H-^{15}N HSQC spectrum of the mini-barnase(modified) which has ten ^{15}N-labeled residues, Ala 11, Tyr 13, Leu 14, Leu 89, Tyr 90, Ser 91, Ser 92, Asp 93, Leu 95, and Tyr 97 (Fig IV-4). Dispersion of the cross-peaks of the labeled residues in the mini-barnase are much larger than that of cross-peaks of Ala, Tyr, Leu, Ser, and Asp in random coil state; the one is 10.2-7.9 ppm in ^{1}H axis and 129-114 ppm in ^{15}N axis and the other is 8.4-8.0 ppm in ^{1}H axis and 123-116 ppm in ^{15}N axis [51,52]. It indicates that the mini-barnase is in neither random coil nor molten globule states, and forms a stable specific conformation.

^{1}H-^{15}N HSQC spectrum of uniformly ^{15}N-labeled barnase has been measured at pH 6.6, 30 °C and the complete assignment of the cross-peaks has been reported [53]. We picked up ten cross-peaks of the corresponding residues to the ^{15}N-labeled residues in the mini-barnase, and compared them with the unassigned cross-peaks in the ^{1}H-^{15}N HSQC spectrum of the mini-barnase. Because of the difference in solvent conditions such as pH and temperature, the two sets of values of chemical shifts in the two measurements are systematically different from each other to some extent. The ^{1}H-^{15}N HSQC spectrum of the mini-barnase was shifted by 0.1 and 4.6 ppm along ^{1}H and ^{15}N axes respectively, and superimposed on that of the picked-up cross-peaks of barnase(Fig. IV-5). Patterns of cross-peak locations in the two spectra are so similar that we can see one-to-one correspondence between the cross-peaks of the mini-barnase and those of barnase. By using this correspondence we could assign all of the ten cross-peaks to the ^{15}N-labeled residues of the mini-barnase(Table IV-3). Chemical shift differences of the corresponding cross-peaks are small as shown in Table IV-3.

If the 3D structure of the mini-barnase is similar to that of the

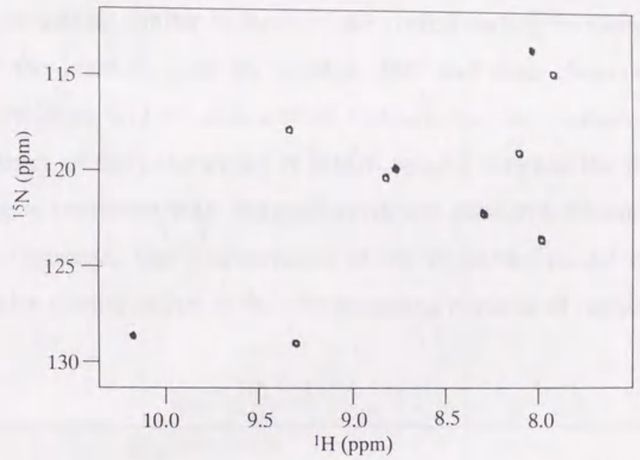

Fig. IV-3. ^{1}H-^{15}N HSQC spectrum of the mini-barnase site-specifically labeled with ^{15}N. Cross-peaks of amide protons and ^{15}Ns of the ten residues, Ala 11 Tyr 13, Leu 14, Leu 89, Tyr 90, Ser 91 ,Ser 92 Asp 93, Leu 95, and Tyr 97 are shown, but not assigned.

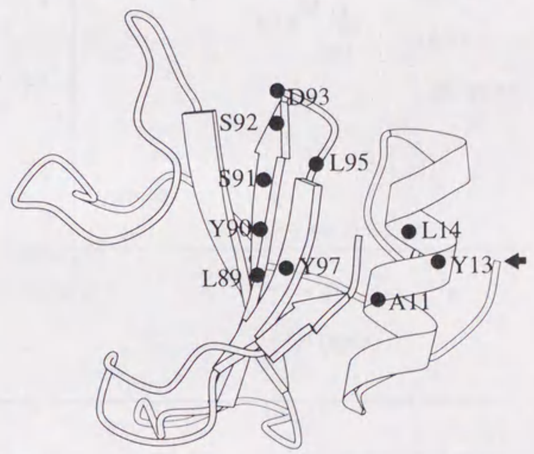

Fig. IV-4. ^{15}N-labeled residues in the conformation of the molecular-modeled mini-barnase which is shown in the direction of the right side of Fig. IV-1 b. ●: ^{15}N-labeled residues. One-letter-codes for amino acid are used. The N-terminus is indicated by a filled arrow.

corresponding portions of barnase, local magnetic fields of residues of the mini-barnase will be similar to those of the corresponding residues of barnase except for the residues near the module M2, and then chemical shifts of respective residues will be also similar between the two proteins. The good correspondence of the cross-peaks of HSQC spectra between the mini-barnase and barnase is consistent with the conformational similarity between the mini-barnase and barnase. The conformation of our molecular model of the mini-barnase is also similar to that of the corresponding portions of barnase. Three of

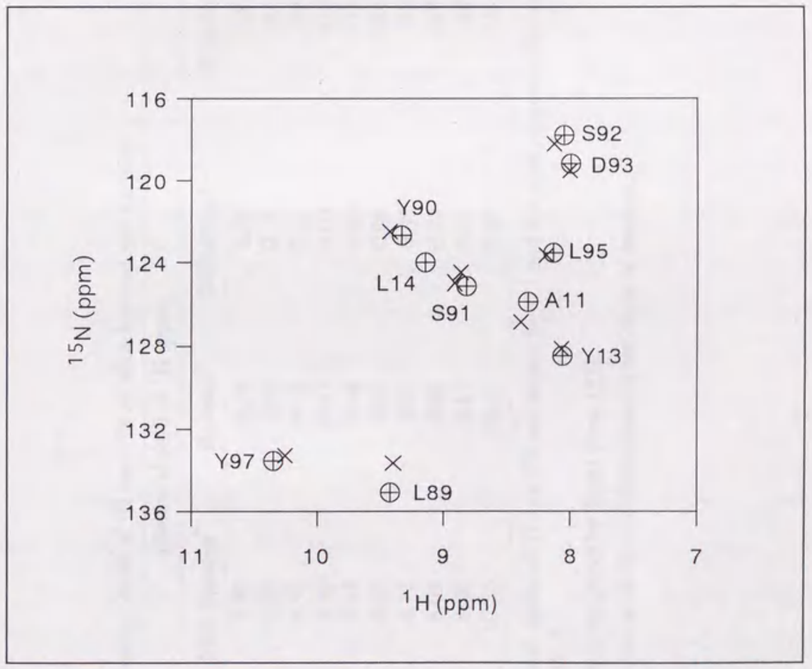

Fig. IV-5. Location of the unassigned cross-peaks of the ^{15}N-labeled residues of the mini-barnase in the ^{1}H-^{15}N HSQC spectrum: ×, and those of assigned cross-peaks of the corresponding residues of barnase in the ^{1}H-^{15}N HSQC spectrum [53]:⊕. The cross-peaks in the ^{1}H-^{15}N HSQC spectrum of the mini-barnase shown in Fig. IV-3 are shifted here by 0.1 and 4.6 ppm along ^{1}H and ^{15}N axes, respectively. One-letter-codes of amino acid are used.

Table IV-3. Chemical Shifts of ^{1}H and ^{15}N of Mini-barnase and Barnase.

	Chemical shift of ^{1}H (ppm)			Chemical shift of ^{15}N (ppm)		
Residue	Mini-barnase*	Barnase†	Difference‡	Mini-barnase*	Barnase†	Difference‡
Ala 11	8.40	8.34	0.06	126.9	125.9	1.0
Tyr 13	8.07	8.06	0.01	128.2	128.5	-0.3
Leu 14	8.92	9.15	-0.23	125.0	124.0	1.0
Leu 89	9.40	9.43	-0.03	133.7	135.1	-1.4
Tyr 90	9.43	9.34	0.09	122.5	122.7	-0.2
Ser 91	8.87	8.83	0.04	124.5	125.1	-0.6
Ser 92	8.13	8.05	0.08	118.2	117.7	0.5
Asp 93	8.01	8.00	0.01	119.5	119.1	0.4
Leu 95	8.20	8.14	0.06	123.7	123.5	0.2
Tyr 97	10.26	10.35	-0.09	133.3	133.5	-0.2

*Values of chemical shifts of ^{1}H and ^{15}N are shifted by 0.1 and 4.6 ppm, respectively, and assigned based on the correspondence shown in Figure IV-5.

†Values are given in the paper by Jones *et al.* [53].

‡Difference of chemical shifts of mini-barnase from those of barnase.

the ten labeled residues, Ala 11, Tyr 13 and Leu 14, and the other seven residues, Leu 89, Tyr 90, Ser 91, Ser 92, Asp 93, Leu 95 and Tyr 97, are located on the buried face of the N-terminal α-helix in module M1 and in the β-hairpin in M5, respectively, in the 3D structure of barnase and the molecular model of mini-barnase(Fig. IV-4). The β-hairpin is the central part of the five-stranded anti-parallel β-sheet, which forms a hydrophobic core with the N-terminal α-helix. The results of our NMR measurements indicate that the mini-barnase forms a stable conformation as we have designed, at least for regions around the hydrophobic core sandwiched the α-helix and the β-sheet.

In process of the sample preparation for NMR measurement, we found the mini-barnase dissolved in water at 1 mM concentration at pH 4.1. For 3D structure determination by NMR, the protein should dissolve at least 1 mM concentration and should form a stable specific conformation. A pH titration study of acid denaturation of the mini-barnase monitored by far-UV CD spectrum showed a conformational transition occurred between pH 3.0 and 3.5, and the CD spectrum at pH 4.1 was the same as that at pH 5.1(unpublished data). This fact indicates that the mini-barnase takes on the same conformation at pH 4.1 as that at pH 5.1 where our NMR measurement was done .

IV-4. Discussion

We designed a new protein by removing the second module from a globular protein, barnase. The designed mini-barnase site-specifically labeled with ^{15}N was chemically synthesized and its solution structure was investigated by 2D NMR. The NMR spectrum indicated that the mini-barnase folded into a stable specific conformation, which appeared to be similar to the designed conformation. The results of our study demonstrate the possibility of our approach to design a new protein by modifying module organization of a present-day protein. In order to prove the success of the design of the mini-

barnase, however, we have to determine its 3D structure and show the determined structure identical or similar to the designed. The mini-barnase dissolved in water at concentration high enough for 3D structure determination by NMR. Determination of its 3D structure is on-going.

In the process of designing the mini-barnase, we succeeded in modeling its conformation, which is mechanically stable. Twenty six of the 28 residues of the second module were removed from barnase. Deletion of such a large fragment from a globular domain is generally thought difficult in molecular modeling. We could, however, make the model easily for the following reasons. First, N- and C-terminal ends of a module are close in space because of the compactness of the module. This characteristic of modules makes easy the connection of two chains left after the removal of a module, if necessary, by shifting the excision points slightly from the module boundaries. Second, conformations of most modules are mechanically stable by themselves (K.Takahashi *et al.*, in press, and [54]; see also chapters II and III). Thus, after the removal of a module from a globular domain, the conformation of the remaining parts is expected to be mechanically stable. These general characteristics of modules may possibly enable us to make a model of another mini-barnase by removing another module. The module M2 is the most independent of other modules [31,49]. We chose the M2 to remove in our first challenge to design of mini-barnase, because the possibility of the success was expected to be the highest. Designs of other mini-barnases and their experimental examinations, however, will provide helpful information to develop the module-based protein design.

Amino acid replacement was done on the two hydrophobic residues on the new surface of the mini-barnase in order to increase the thermodynamic stability and the solubility of the folded mini-barnase, being based on the theoretical prediction method. The experimental results of the conformational

stability and the solubility of the designed protein indicate the validity of the method of the identification of residues to replace.

Preceding our study, De Sanctis *et al.* [55,56] made a mini-myoglobin of 108 amino acid residues, which is a peptide excised from myoglobin and closely corresponds to the peptide encoded by the central exon but contains an additional 30 residues at the carboxy-terminal end. The peptide binds heme stoichiometrically, forms a helical structure, and exhibits the reversible oxygen-binding function of myoglobin itself. Pessi *et al.* [57] also designed a mini-protein of 61 amino acid residues named minibody using the central portion of the heavy chain variable domain of an immunoglobulin as a template which constitutes a compact sub-domain consisting of a β-sandwich of two triple-stranded β-sheets and containing the hypervariable loops H1 and H2. Hydrophobic residues buried in the parent molecule but exposed in the minibody were suitably replaced and a metal-binding site was engineered by introducing one histidine in H1 and two in H2. The minibody exhibited all-β structure features in CD spectra, cooperative transition in urea-induced unfolding, and Zn-binding activity, showing success of the design. Design of mini-proteins itself will become an important technology in protein engineering, because the small size of designed proteins makes easy theoretical and experimental treatments such as molecular dynamics simulation and structure determination by NMR.

Design of globular proteins by exchanging modules or exons between homologous proteins has been studied by a few groups. Kumagai *et al.* [58] succeeded in functional conversion of α-lactalbumin and the c-type lysozyme by exon exchange. These two proteins are functionally different but structurally homologous. Their genes have the same exon-intron organization. Exchange of the second exon of α-lactalbumin with that of lysozyme produced chimera α-lactalbumin with lysozyme activity, whereas only replacement of ten amino

acids directly involved in substrate-binding did not make α-lactalbumin acquire lysozyme activity. Yaoi *et al.* [59] also successfully converted the coenzyme specificity of isocitrate dehydrogenase (ICDH) from NADP-dependent to NAD-dependent by replacing a module involved in the coenzyme binding site with the corresponding module of the homologous NAD-dependent enzyme 3-isopropylmalate dehydrogenase (IPMDH). Replacement of a few residues interacting with 2'-phosphate group of NADP did not result in the conversion of the coenzyme specificity and seems to have unstabilized the structure of ICDH. The module-replaced ICDH was as stable as the original wild type ICDH. Wakasugi *et al.* replaced the fourth module of β-subunit of hemoglobin and myoglobin by the fourth module of α-subunit of hemoglobin. They found the chimeric globin associated with the β-subunit of hemoglobin through interactions between the fourth module of α-subunit in the chimeras and the β-subunit [60,61]. These studies demonstrate that modules are functional units as well as structural units of proteins, and the possibility of design of functionally chimeric proteins by exchanging modules. The module exchange is, however, limited within homologous proteins. To develop the protein design method of creating a new protein, a new assembly of modules even from evolutionary unrelated proteins, we have to understand physicochemical bases for the conformational stability of the module assembly. Though studies of mini-proteins designed by module excision will also elucidate roles of modules in protein functions, we place our challenge of making the mini-barnase at the first step in course of elucidation of the physicochemical bases for conformations of module assemblies.

Fersht's group has been studying folding process and conformational stability of barnase mainly by using systematic amino acid replacement, and has obtained valuable information [62-64]. They have also studied truncated barnases of which N-terminal 22 and 36 residues were cut off [65,66]. The

truncated barnases did not appear to form stable conformations in the isolated state. Comparison between results of their studies and those of our study of the mini-barnase will give a new insight into folding process and conformational stability of barnase. Especially elucidation of roles of modules in protein folding and stability may provide a hint on understanding relation between protein folding and protein evolution. Results of the comparison will be reported elsewhere.

IV-5. References

1. Shao, Z.X.,Arnold, F.H. Engineering new functions and altering existing functions. Curr. Opin. Struct. Biol. 6:513-518, 1996.

2. Shaw, A.,Bott, R. Engineering enzymes for stability. Curr. Opin. Struct. Biol. 6:546-550, 1996.

3. Hedstrom, L. Engineering for redesign. Curr. Opin. Struct. Biol. 4:608-611, 1994.

4. Fersht, A.,Winter, G. Protein engineering. Trends Biochem. Sci. 17:292-294, 1992.

5. Betz, S.F., Bryson, J.W.,DeGrado, W.F. Native-like and structurally characterized designed α-helical bundles. Curr. Opin. Struct. Biol. 5:457-463, 1995.

6. O'Shea, E.K., Lumb, K.J.,Kim, P.S. Peptide 'Velcro': design of a heterodimeric coiled coil. Curr. Biol. 3:658-667, 1993.

7. Raleigh, D.P., Betz, S.F.,DeGrado, W.F. A *de novo* designed protein mimics the native state of natural proteins. J. Am. Chem. Soc. 117:7558-7559, 1995.

8. Betz, S.F.,DeGrado, W.F. Controlling topology and native-like behavior of *de novo*-designed peptides: design and characterization of antiparallel four-stranded coiled coils. Biochemistry 35:6955-6962, 1996.

9. Karle, I.L., Awasthi, S.K.,Balaram, P. A designed β-hairpin peptide in crystals. Proc. Natl. Acad. Sci. USA 93:8189-8193, 1996.

10. Ramirez-Alvarado, M., Blanco, F.J.,Serrano, L. *De novo* design and structural analysis of a model β-hairpin peptide system. Nature Struct. Biol. 3:604-612, 1996.

11. Betz, S.F., Raleigh, D.P.,DeGrado, W.F. *De novo* protein: from molten globules to native-like states. Curr. Opin. Struct. Biol. 3:601-610, 1993.

12. Tanaka, T., Kuroda, Y., Kimura, H., Kidokoro, S.,Nakamura, H. Cooperative deformation of a *de novo* designed protein. Protein Eng. 7:969-976, 1994.

13. Houbrechts, A., Moreau, B., Abagyan, R., Mainfroid, V., Preaux, G., Lamproye, A., Poncin, A., Goormaghtigh, E., Ruysschaert, J.M., Martial, J.A.,et al, Second-generation octarellins: two new *de novo* $(\beta/\alpha)_8$ polypeptides designed for investigating the influence of β-residue packing on the α/β-barrel structure stability. Protein Eng. 8:249-259, 1995.

14. Smith, D.D., Pratt, K.A., Sumner, I.G.,Henneke, C.M. Greek key jellyroll protein motif design: expression and characterization of a first-generation molecule. Protein Eng. 8:13-20, 1995.

15. Gō, M. Correlation of DNA exonic regions with protein structural units in haemoglobin. Nature 291:90-92, 1981.

16. Gō, M. Modular structural units, exons, and function in chicken lysozyme. Proc. Natl. Acad. Sci. USA 80:1964-1968, 1983.

17. Jensen, E.O., Paludan, K., Hyldig-Nielsen, J.J., Jorgensen, P.,Marcker, K. The structure of a chromosomal leghaemoglobin gene from soybean. Nature 291:677-679, 1981.

18. Blake, C.C. Exons and the structure, function and evolution of haemoglobin. Nature 291:616, 1981.

19. Isaacs, N.W., Machin, K.J.,Masakuni, M. Three-dimensional structure of goose-type lysozyme from the egg white of the Australian black swan, *Cygnus atratus*. Aust. J. Biol. Sci. 38:13-22, 1985.

20. Gō, M.,Nosaka, M. Protein architecture and the origin of introns. Cold spring Harb. Symp. Quant. Biol. 52:915-924, 1987.

21. Gilbert, W., Marchionni, M.,McKnight, G. On the antiquity of introns. Cell 46:151-153, 1986.

22. Gilbert, W.,Glynias, M. On the ancient nature of introns. Gene 135:137-144, 1993.

23. Tittiger, C., Whyard, S.,Walker, V.K. A novel intron site in the triosephosphate isomerase gene from the mosquito *Culex tarsalis*. Nature 361:470-472, 1993.

24. Gō, M. Protein structures and split genes. Adv. Biophys. 19:91-131, 1985.

25. Gilbert, W. Why genes in pieces? Nature 271:501-500, 1978.

26. Straus, D.,Gilbert, W. Genetic engineering in the Precambrian: structure of the chicken triosephosphate isomerase gene. Mol. Cell Biol. 5:3497-3506, 1985.

27. Blake, C.C.F. Exons and the evolution of proteins. Int. Rev. Cytol. 93:149-185, 1985.

28. Gilbert, W. Genes-in-pieces revisited. Science 228:823-824, 1985.

29. Nishimura, S.,Nomura, M. Ribonuclease of *bacillus subtilis*. J. Biochem. 46:161-167, 1959.

30. Hartley, R.W.,Barker, E.A. Amino-acid sequence of extracellular ribonuclease (barnase) of *Bacillus amyloliquefaciens*. Nature New Biol. 235:15-16, 1972.

31. Noguti, T., Sakakibara, H.,Gō, M. Localization of hydrogen-bonds within modules in barnase. Proteins 16:357-363, 1993.

32. Ikura, T., Gō, N., Kohda, D., Inagaki, F., Yanagawa, H., Kawabata, M., Kawabata, S., Iwanaga, S., Noguti, T.,Gō, M. Secondary structural features of modules M2 and M3 of barnase in solution by NMR experiment and distance geometry calculation. Proteins 16:341-356, 1993.

33. Yoshida, K., Shibata, T., Masai, J., Sato, K., Noguti, T., Gō, M.,Yanagawa, H. Protein anatomy: spontaneous formation of filamentous helical structures from the N-terminal module of barnase. Biochemistry 32:2162-2166, 1993.

34. Mauguen, Y., Hartley, R.W., Dodson, E.J., Dodson, G.G., Bricogne, G., Chothia, C.,Jack, A. Molecular structure of a new family of ribonucleases. Nature 297:162-164, 1982.

35. Baudet, S.,Janin, J. Crystal structure of a barnase-d(GpC) complex at 1.9 Å resolution. J. Mol. Biol. 219:123-132, 1991.

36. Guillet, V., Lapthorn, A., Hartley, R.W.,Mauguen, Y. Recognition between a bacterial ribonuclease, barnase, and its natural inhibitor, barster. Structure 1:165-177, 1993.

37. Bycroft, M., Ludvigsen, S., Fersht, A.R.,Poulsen, F.M. Determination of the three-dimensional solution structure of barnase using nuclear magnetic resonance spectroscopy. Biochemistry 30:8697-8701, 1991.

38. Mayo, S.L., Olafson, B.D.,Goddard, W.A.III. Dreiding: a generic force field for molecular simulations. J. Phys. Chem. 94:8897-8909, 1990.

39. Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.E.,Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T.,Tasumi, M. The Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol. 112:535-542, 1977.

40. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A.,Haak, J.R. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81:3684-3690, 1984.

41. Oobatake, M.,Ooi, T. Hydration and heat stability effects on protein unfolding. Prog. Biophys. Mol. Biol. 59:237-284, 1993.

42. Ooi, T.,Oobatake, M. Prediction of the thermodynamics of protein unfolding: the helix-

coil transition of poly(L-alanine). Proc. Natl. Acad. Sci. USA 88:2859-2863, 1991.

43. Ooi, T., Oobatake, M., Nemethy, G.,Scheraga, H.A. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. Proc. Natl. Acad. Sci. USA 84:3086-3090, 1987.

44. Ooi, T.,Oobatake, M. Intermolecular interactions between protein and other molecules including hydration effects. J. Biochem. 104:440-444, 1988.

45. Shrake, A.,Rupley, J.A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. J. Mol. Biol. 79:351-371, 1973.

46. Gō, M.,Miyazawa, S. Relationship between mutability, polarity and exteriority of amino acid residues in protein evolution. Int. J. Peptide Protein Res. 15:211-224, 1980.

47. Hojo, H.,Aimoto, S. Synthesis of barnase site-specifically labelled with two ^{13}C atoms using partially protected peptide thioester building blocks. Bull. Chem. Soc. Jpn. 66:3004-3008, 1993.

48. Boyd, J., Soffe, N., John, B., Plant, D.,Hurd, R. The generation of phase-sensitive 2D ^{15}N-^{1}H spectra using gradient pulses for coherence-transfer-pathway selection. J. Magn. Reson. 98:660-664, 1992.

49. Noguti, T. and Gō, M. Modules of barnase:the physicochemical basis for their structures. In: "Tracing Biological Evolution in Protein and Gene Structures." Gō, M., Schimmel, P. (eds.). Amsterdam: Elsevier, 1995:161-174.

50. Dauber-Osguthorpe, P., Roberts, V.A., Osguthorpe, D.J., Wolff, J., Genest, M.,Hagler, A.T. Structure and energetics of ligand binding to proteins: *E. coli* dihydrofolate reductase-trimethoprim, a drug-receptor system. Proteins 4:31-47, 1988.

51. Bundi, A., Wüthrich, K. ^{1}H-NMR parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. Biopolymers 18:285-297, 1979.

52. Wishart, D.S., Sykes, B.D.,Richards, F.M. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. J. Mol. Biol. 222:311-333, 1991.

53. Jones, D.N., Bycroft, M., Lubienski, M.J.,Fersht, A.R. Identification of the barstar binding site of barnase by NMR spectroscopy and hydrogen-deuterium exchange. FEBS Lett. 331:165-172, 1993.

54. Takahashi, K., Gō, M. and Noguti, T. Mechanical stability of protein modules determined by molecular dynamics simulations. In: "Tracing Biological Evolution in Protein and Gene Structures." Gō, M., Schimmel, P. (eds.). Amsterdam: Elsevier, 1995:175-185.

55. De Sanctis, G., Falcioni, G., Giardina, B., Ascoli, F.,Brunori, M. Mini-myoglobin. The

structural significance of haem-ligand interactions. J. Mol. Biol. 200:725-733, 1988.

56. De Sanctis, G., Falcioni, G., Giardina, B., Ascoli, F.,Brunori, M. Mini-myoglobin: preparation and reaction with oxygen and carbon monoxide. J. Mol. Biol. 188:73-76, 1986.

57. Pessi, A., Bianchi, E., Crameri, A., Venturini, S., Tramontano, A.,Sollazzo, M. A designed metal-binding protein with a novel fold. Nature 362:367-369, 1993.

58. Kumagai, I., Takeda, S.,Miura, K. Functional conversion of the homologous proteins α-lactalbumin and lysozyme by exon exchange. Proc. Natl. Acad. Sci. USA 89:5887-5891, 1992.

59. Yaoi, T., Miyazaki, K., Oshima, T., Komukai, Y.,Gō, M. Conversion of the coenzyme specificity of isocitrate dehydrogenase by module replacement. J. Biochem. 119:1014-1018, 1996.

60. Wakasugi, K., Ishimori, K., Imai, K., Wada, Y.,Morishima, I. "Module" substitution in hemoglobin subunits. Preparation and characterization of a "chimera βα-subunit". J. Biol. Chem. 269:18750-18756, 1994.

61. Wakasugi, K., Ishimori, K. and Morishima, I. Module substitution in globins: preparation and association characteristics of chimeric hemoglobin subunits and myoglobin. In: "Tracing Biological Evolution in Protein and Gene Structures." Gō, M., Schimmel, P. (eds.). Amsterdam: Elsevier, 1995:283-295.

62. Matouschek, A., Kellis, J.T.,Jr., Serrano, L., Bycroft, M.,Fersht, A.R. Transient folding intermediates characterized by protein engineering. Nature 346:440-445, 1990.

63. Matouschek, A., Kellis, J.T.,Jr., Serrano, L.,Fersht, A.R. Mapping the transition state and pathway of protein folding by protein engineering. Nature 340:122-126, 1989.

64. Fersht, A.R. The sixth Datta Lecture. Protein folding and stability: the pathway of folding of barnase. FEBS Lett. 325:5-16, 1993.

65. Kippen, A.D., Sancho, J.,Fersht, A.R. Folding of barnase in parts. Biochemistry 33:3778-3786, 1994.

66. Sancho, J.,Fersht, A.R. Dissection of an enzyme by protein engineering. The N and C-terminal fragments of barnase form a native-like complex with restored enzymic activity. J. Mol. Biol. 224:741-747, 1992.

CONCLUSION

Molecular dynamics simulations were performed on the six modules of barnase starting with their native conformations, for one nanosecond in vacuum and for 150 picoseconds in water. Five of the six excised modules retained native-like, compact conformations during these simulations both in vacuum and in water. Similar simulations were also carried out on pseudo modules as a control study. Contrary to modules, all pseudo modules were largely deformed in 100 picoseconds. These results show that almost all modules of barnase are mechanically stable (*i.e.* the balance of interatomic forces is accomplished within individual module), whereas all pseudo modules are mechanically unstable.

The mechanical stability of modules might be rather unexpected results, because module boundaries are often located on α-helices or β-sheets; these structural elements are stabilized by hydrogen bonds. The structural stability of excised modules suggests that conformations of modules are less dependent of their molecular environments and thus the conformations are retained when the modules are brought into a new molecular environment through a new combination of modules. This autonomous nature of modules is suitable for recruited units. Therefore the mechanical stability of modules gives physico-chemical bases for modules to be original building blocks of proteins.

It has been also shown from molecular dynamics study, that most of the modules of barnase are mechanically stable without ionic charge and hydrogen bond interactions (chapter III). This demonstrates that the modules are stabilized by sum of many interactions that are weak per atomic pair such as *van der Waals* and non-ionic partial charge interactions. This large number of atomic

pairs interacting in a module is due to compact conformation of modules. Therefore it has been concluded that compactness of modules is essential for their mechanical stability. Compact modules might have been selected as useful evolutionary units encoded by exons due to their conformational stability necessary for building blocks.

It has been indicated that a mini-barnase designed by removing the second module M2 from barnase, folds into a stable conformation similar to that of barnase (chapter IV). The structure of the site-specifically ^{15}N-labeled mini-barnase was examined by 2D NMR measurement. It was found that both ^{1}H and ^{15}N chemical shifts of the mini-barnase corresponded well to those of barnase. This indicates that the structures of both proteins are similar except for the removed module. Because deletion of the module did not largely alter the parent protein's structure, there should be little mutual dependence among the conformations of modules. The mini-barnase without module M2 might have existed as an evolutionary intermediate state. The result that the stable mini-protein was created by module-based approach has opened the possibility of designing novel proteins by new combinations of modules. The present study also gives experimental support to the view that proteins were created by combination of modules during evolution.

ACKNOWLEDGEMENTS

I wish to express my great appreciation to Professor Mitiko Gō for her instruction, encouragement and support in the present work. I also wish to express my thanks to Dr. Tosiyuki Noguti for his support and valuable discussions in this study. I am very grateful to thank Drs. Tadayasu Ohkubo, Hironobu Hojo and Saburo Aimoto for their cooperation in the present experimental study, Drs. Teikichi Ikura and Kunihiro Kuwajima for their guidance in CD measurement, Drs. Rieko Ishima and Kuniaki Nagayama for their instruction in NMR measurement, Drs. Yasuo Sugiyama and Yasuo Mukohata, Drs. Akira Mizoguchi and Yasuyoshi Nishida, Drs. Kiyoshi Yamauchi and Masayoshi Kinoshita, and Dr. Tadao Kondo and members of Chemical Instrument Center of Nagoya University, for their kindness of allowing me to use their experimental equipments and useful advice. I would like to thank Dr. Kaoru Fukami-Kobayashi, Mr. Kei Yura and all other colleagues for their help and useful discussions.

副論文

1. Mechanical stability of compact modules of barnase.
K. Takahashi, M. Oohashi, T. Noguti & M. Gō.
FEBS Lett. (in press)
（バルナーゼのコンパクトなモジュールの力学的安定性）

2. Mechanical stability of protein modules determined by molecular dynamics simulations.
K.Takahashi, M. Gō & T. Noguti
In: "Tracing biological evolution in protein and gene structures." M. Gō & P. Schimmel (eds.)
Amsterdam: Elsevier, 1995:175-185.
（分子動力学計算により決定されたタンパク質モジュールの力学的安定性）