



A study for elucidation of architectural principle of protein structure

Araki, Mitsugu

(Degree)

博士 (理学)

(Date of Degree)

2007-03-25

(Date of Publication)

2012-05-02

(Resource Type)

doctoral thesis

(Report Number)

甲4016

(URL)

<https://hdl.handle.net/20.500.14094/D1004016>

※ 当コンテンツは神戸大学の学術成果です。無断複製・不正使用等を禁じます。著作権法で認められている範囲内で、適切にご利用ください。



Doctoral dissertation

A study for elucidation of architectural principle of protein structure

2007, February

Mitsugu Araki

Graduate School of Science and Technology, Kobe University

Doctoral dissertation

A study for elucidation of architectural principle of protein structure

(タンパク質立体構造構築原理の解明に向けた研究)

2007, February

Mitsugu Araki

Graduate School of Science and Technology, Kobe University, Nada, Kobe, Japan

Preface

For elucidation of architectural principle of protein structure, the relationship between protein sequence and the tertiary structure has been studied from various perspectives. Previously, determining factors related to protein sequence of the structural stabilizing mechanism have been suggested, mostly, by studies of stability and folding kinetics of natural proteins and the mutagenesis studies ¹⁻³. Recently, computational protein designs have advanced to provide new insights into the determinants of protein structure, stability, and folding ⁴. Computational methods for identifying amino acid sequences compatible with a known target structure have allowed redesign of naturally occurring proteins ⁵⁻⁷. On the other hand, proteins with novel structures have been also created by methods of computational design ^{8,9}. These successful computational designs to create novel protein sequences and structures not only suggest that the potential function guiding the design process captures much of the important physical chemistry, but make it possible to elucidate the properties of natural proteins, which have been selected by various evolution pressures, by comparing the properties of artificially designed counterparts. Nevertheless, all of unique aspects of protein structure would not be exposed by above-described studies. How has nature primarily created folded proteins in case the computational design methods rarely find a protein having a well-packing structure? There should be some kind of trick in forming protein structure. For example, since a protein molecule is a compound of amino acids linked by peptide bonds in a linear sequence, a merit in forming protein structure might be hidden in a protein property in which one polypeptide is linked to another one by a peptide bond. Thus, we examined that how addition of a peptide fragment to the C-terminus of a protein affects the added protein or the whole protein structure. These experimental results would not

only suggest a factor necessary to induce a whole protein fold, but explore a minimal principle that determines the protein structure.

1. Carlsson, U. & Jonsson, B. H. Folding of beta-sheet proteins. *Curr Opin Struct Biol* 5, 482-7 (1995).
2. Chakrabartty, A. & Baldwin, R. L. Stability of alpha-helices. *Adv Protein Chem* 46, 141-76 (1995).
3. Jackson, S. E. How do small single-domain proteins fold? *Fold Des* 3, R81-91 (1998).
4. Kuhlman, B. & Baker, D. Exploring folding free energy landscapes using computational protein design. *Curr Opin Struct Biol* 14, 89-95 (2004).
5. Desjarlais, J. R. & Handel, T. M. De novo design of the hydrophobic cores of proteins. *Protein Sci* 4, 2006-18 (1995).
6. Dahiyat, B. I. & Mayo, S. L. De novo protein design: fully automated sequence selection. *Science* 278, 82-7 (1997).
7. Ponder, J. W. & Richards, F. M. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J Mol Biol* 193, 775-91 (1987).
8. Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T. & Kim, P. S. High-resolution protein design with backbone freedom. *Science* 282, 1462-7 (1998).
9. Kuhlman, B. et al. Design of a novel globular protein fold with atomic-level accuracy. *Science* 302, 1364-8 (2003).

Contents

Chapter 1.

Transformation of an α -helix peptide into a β -hairpin induced by addition of a fragment results in creation of a coexisting state

Abstract2

Introduction3

Materials and methods5

Results7

Discussion13

References17

Chapter 2.

Protein segment that drastically decreases the solubility induces formation of the whole protein structure

Abstract27

Introduction28

Materials and methods31

Results34

Discussion39

References46

Acknowledgements.....58

Publication Lists.....58

Chapter 1.

Transformation of an α -helix peptide into a β -hairpin induced by addition of a fragment results in creation of a coexisting state

Abbreviations: TP:Target Peptide, DP:Designed Peptide, NOE: nuclear overhauser effect, NMR: nuclear magnetic resonance, CD: circular dichroism.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank (PDB ID codes 2DX2 for TP, 2DX3 and 2DX4 for α -helix and β -hairpin conformations for DP5, respectively.)

Abstract

Intrinsic rules of determining the tertiary structure of a protein have been unknown partly because physicochemical factors that contribute to stabilization of a protein structure cannot be represented as a linear combination of local interactions. To clarify the rules on the nonlinear term caused by nonlocal interaction in a protein, we tried to transform a peptide that has a fully helical structure (“Target Peptide” or TP) into a peptide that has a β -hairpin structure (“Designed Peptide” or DP) by adding seven residues to the C terminus of TP. According to analyses of nuclear magnetic resonance measurements, while the β -hairpin structure is stabilized in some DPs, it is evident that the helical structure observed in TP is also persistent and even extended throughout the length of the molecule. As a result, we have produced a peptide molecule that contains both the α -helix and β -hairpin conformation at an almost equally populated level. The helical structures contained in these DPs were more stable than the helix in TP, suggesting that stabilizing one conformation does not result in destabilizing the other conformation. These DPs can thus be regarded as an isolated peptide version of the chameleon sequence, which has capability of changing the secondary structure depending on the context of the surrounding environment in a protein structure. The fact that the transformation of one secondary structure caused stabilization of both the original and induced structure would shed light on the mechanism of protein folding.

Introduction

To understand the stabilizing mechanism of a protein structure, numerous studies have focused on the secondary structure, which is the backbone of a protein structure and mainly contains the α -helix, β -strand, and loop. In general, α -helix or β -sheet forming propensity in a protein depends on both intrinsic capability of individual amino acids to form the local secondary structure and ability to interact with the surrounding tertiary structure. It was shown that some correlations exist between the α helical propensities obtained from host-guest experiments ¹⁻⁶ and statistical frequencies of α helix ⁷, indicating that the intrinsic ability of each amino acid largely contributes to the α helix formation. On the other hand, there is little correlation between statistical frequencies of β sheet ⁷ and β sheet propensity, even at solvent-exposed position ⁸⁻¹⁰. It is then concluded that ability to interact with surrounding tertiary structure, rather than the intrinsic ability of each amino acid, largely contributes to the β sheet formation. In addition, Regan et al. transmuted a primarily β sheet protein into a four-helix bundle protein while retaining 50% of the original sequence by taking account of the tertiary structure ¹¹. Their experimental result suggests that less than 50% of amino acid residues in the protein sequence can determine the secondary structure while the remaining half is irrelevant to the structural formation, suggesting that helix or sheet propensities are not the sole source of formation of the structure. Thus, we still have long way to understand the concrete rules of determining the secondary and tertiary structure in a protein.

Several problems prevent us from understanding the rules of formation of the secondary structure. First, the number of atoms that make up a protein molecule is very large and they interact with one another. Second, protein molecules exist in aqueous

solution with marginal stability at a cost of entropy of both the protein and water molecules. As a result, a protein fragment does not generally have a structure that corresponds to the secondary structure observed in the native protein. In other words, the free energy of a protein unfolding is not equal to the sum of free energies of unfolding of the isolated fragments. To fill up the difference, an adjusting term is necessary and this can be called as an interaction term, since it is caused by the interaction between the fragments consisting of secondary structural units. Therefore, we think that clarification of this term would result in further understanding of the determining factor of the secondary as well as tertiary structure in proteins.

Here, as a experimental strategy to clarify the rules on this interaction term, we tried to transform a peptide (defined as “Targeted Peptide” or “TP”) that has a fully helical structure into a peptide (defined as “Designed Peptide” or “DP”) that has a fully β -hairpin structure consisting of two antiparallel β -strands by adding several residues to the C terminus of the original peptide, and not by substitution of amino acids commonly employed. After appropriate elongated peptide sequences were established and synthesized, overall structure and stability of the helical portion of DPs were compared to those of TP, and then we characterized the interaction term of the helical structure induced by the addition of several residues to the C-terminus. What is particularly interesting and apparently contradicting result is that stabilization of the newly introduced β -hairpin caused *stabilization* of the original α -helix. We will try to explain how and why this has occurred.

Materials and Methods

Peptide synthesis and purification. Peptides were synthesized by Pioneer Peptide Synthesis System (PerSeptive Biosystems, CA, USA) using Fmoc solid-phase chemistry and were cleaved from the resin with a solution containing 82.5% (vol/vol) trifluoroacetic acid (TFA), 5% H₂O, 5% thioanisole, 2.5% 1,2-ethanedithiol, and 0.8M phenol. Individual peptides were purified by reverse-phase HPLC (acetonitrile/H₂O/0.1%TFA). Peptide identity was confirmed by laser desorption time of flight mass spectrometry, AXIMA-CFR (SHIMADZU, Kyoto, Japan).

Circular Dichroism (CD) measurements. Spectra were acquired at 10°C on a Jasco J-720 CD spectropolarimeter with a 0.5mm pathlength cuvette on peptide samples of 0.1-0.2mM concentration. A buffer containing 10mM acetic acid, 3mM NaOH in 90% H₂O and 10% D₂O at pH 4.5 was basically used. For urea titrations, a peptide stock solution in this buffer was mixed with a stock solution of 8M urea. Peptide concentrations were determined by absorbance measurements as described ¹².

NMR spectroscopy. NMR spectra were performed on a Bruker DMX-750 spectrometer at 10°C on peptide samples of 1mM concentration. Pulsed-field gradient NMR spectra were acquired at 10°C or 20°C on peptide samples of 0.4-1.5mM concentration. 10mM acetic acid, 3mM NaOH in 90% H₂O and 10% D₂O or in 99.9% D₂O at pH 4.5 were used as a buffer. All chemical shifts were referenced to the sodium salt of trimethylsilylpropionate (TSP). Pulsed-field gradient NMR spectroscopy, double quantum filtered correlation spectroscopy (2QF COSY), total correlation spectroscopy (mixing time 80ms), rotating frame Overhauser effect (ROE) spectroscopy (mixing time

250, 300ms), and nuclear Overhauser effect (NOE) spectroscopy (mixing time 50, 100, 150, 200, 250, 300, and 350ms) experiments were performed and water suppression was achieved by selective presaturation or field-gradient pulses¹³. The proton resonances were assigned by the sequential assignment procedure¹⁴.

Assessment of peptide self-association. Extent of self-association of the peptides was evaluated by the following methods. First, we confirmed that line widths and chemical shifts of all signals observed in one-dimensional ¹H NMR spectra were identical for solutions at 1mM-0.04mM peptide concentration, suggesting that the peptides keep the monomeric state up to 1mM. Second, translational diffusion coefficients (D_{pep}) of the peptides were obtained at 20°C using pulsed-field gradient NMR spectroscopy as described^{15,16}. As a result of transformation of D_{pep} to the hydrodynamic radii (R_h^{pep}) using a reference molecule in the peptide solution, R_h^{pep} of TP, DP1, DP2, DP3, DP4, and DP5 became 8.9 ± 0.6 , 11.3 ± 0.5 , 11.9 ± 0.5 , 11.1 ± 0.5 , 11.5 ± 0.7 , and 10.3 ± 0.6 (Å), respectively¹⁷. On the other hand, theoretical hydrodynamic radii for the monomer ($R_h^{\text{pro/monomer}}$), which were calculated according to the equation for native folded proteins, of TP, and DPs became 9.6 ± 2.7 , and 11.1 ± 3.3 (Å), respectively. In a similar way, theoretical hydrodynamic radii for the dimer ($R_h^{\text{pro/dimer}}$) of TP and DPs became 11.8 ± 3.5 , and 14.0 ± 3.8 (Å), respectively¹⁷. In addition, theoretical hydrodynamic radii for the monomer of random polypeptide chains ($R_h^{\text{random/monomer}}$), which were calculated based on the equation for random polypeptide chains, of TP and DPs become 8.9 ± 4.6 , and 11.8 ± 6.3 (Å), respectively¹⁷. It is thus shown that R_h^{pep} values are close to $R_h^{\text{pro/monomer}}$ or $R_h^{\text{random/monomer}}$ than to $R_h^{\text{pro/dimer}}$, indicating that each peptide takes a monomeric form throughout all the experimental

conditions.

Structure calculations. Distance restraints were obtained by converting integrated NOE peak intensities into distance upper limits, using the macro CALIBA in DYANA¹⁸. Standard pseudo atom distances were used when they were needed. ϕ angles were restricted to $-65 \pm 30^\circ$ for measured $^3J_{\text{NH}\alpha}$ values below 6Hz for TP. No ϕ angle restraints were used for DPs because their measured $^3J_{\text{NH}\alpha}$ values would be averages of multiple states. With a cutoff of 0.2Å for upper bound NOE violations, a total of 50 structures was generated by using DYANA and the 10 lowest energy structures were selected to represent three-dimensional structures.

Results

Sequence design and structure of Targeted Peptide. Among several candidates for TP, we chose a peptide corresponding to residues 101-111 of human α -lactalbumin (α lac:101-111: IDYWLAHKALA), which is known to form an α -helix at low pH^{19,20}. To suppress potential pH sensitivity caused by dissociation of a proton, Asp2 was replaced with Asn beforehand. Then by taking account of the fact that a turn sequence plays an important role in β -hairpin folding^{21,22}, the residues 8-11 (KALA) in the C terminus was replaced with a sequence AKAG, which reportedly has an ability to form a stable 4:4 turn sequence in some short peptides²³, resulting in the sequence of TP (INYWLAHAKAG). We determined the three-dimensional structure of TP by NMR. A total of 155 distance restraints calculated from assigned proton NOEs and four backbone dihedral angle restraints that were derived from DQF-COSY spectra was included in the structure calculations. As a result of the structure calculations, it is

shown that the backbone of residues 3-6 adopts 3_{10} helix conformation (Fig. 1A B, Table2), as observed in α lac:101-111²⁰. The side chains of Y3, W4, and H7 of TP are similarly interacted one another, suggesting that these residues stabilize the 3_{10} helix conformation.

A strategy to design β -hairpin structure and structural analysis of Designed Peptides. We attempted to add extra residues to the C terminus of TP, which forms 3_{10} helix structure, anticipating that the peptide might be transformed into a β -hairpin structure throughout the length of the molecule. Since residues 8-11 (AKAG) of the C terminus of TP are supposed to form the 4:4 turn, the number of additional residues, assuming the second β -strand should be aligned in relation to the first strand, was determined to be seven (X_1 - X_7 in Fig. 2A.). Sequence in the extra region was determined by taking into account the solubility of the amino acids, the frequency of two amino acid pairing within antiparallel β -sheet²⁴, and the aromatic-aromatic (or imidazole-aromatic) cross-strand pairing that contributes to the stabilization of β -hairpin structure through π - π interactions^{22,25-27}. Fig.2B shows the resulting sequences of DPs thus constructed, which contain various degrees of these interactions. It is to be noted that the number of aromatic-aromatic cross-strand pairs of DP1, DP2, DP3-4 and DP5 are 0, 1, 2, 3, respectively. DP2, DP4 and DP5 have a His7-Tyr12 cross-strand pair that is adjacent to the turn region. DP3 and DP5 have a Trp4-His15 cross-strand pair. DP3, DP4, and DP5 have a Tyr3-Trp16 cross-strand pair. Additionally, for comparison of the structural stabilities, we prepared a fragment equivalent to X_1 - X_7 of DP5 (named “DP5¹²⁻¹⁸”).

We found that in the NOESY spectra of some of the successfully designed

peptides, while NOEs that were consistent with the β -hairpin structure were observed, NOEs that indicated the helical structure were also observed. Thus we divided the NOEs for each peptide into three groups: <1> NOEs consistent solely with helical structure: NOEs of $d_{NN}(i, i+1)$, $d_{NN}(i, i+2)$, $d_{\alpha N}(i, i+2)$, and those observed between residues i and $i+3$, i and $i+4$, and side chains of residues i and $i+1$. <2> NOEs consistent solely with a β -hairpin structure: interstrand NOEs consistent with expected β -hairpin conformation, intrastrand NOEs observed between side chains of residues i and $i+2$. <3> Other NOEs consistent with both the helical and β -hairpin structure: intra-residue NOEs, NOEs between backbone and side chain or backbone of residues i and $i+1$, and hydrogens in A8-G11 whose expected β -turn has similar NOE pattern to that of helix. Correlation between the number of NOEs and the residual numbers for the cases <1> and <2> are shown in Fig.3A and Fig.3B, respectively. The spectra of DP1 had NOEs consistent solely with helical structure observed in TP and did not have any NOEs consistent with the expected β -hairpin conformation. The spectra of DP3 showed some intrastrand NOEs without interstrand NOEs and had as much NOEs consistent with α -helix as that of TP. These results indicate that DP1 and DP3 have only helical structure resembling that of TP. In the spectra of DP2, DP4, and DP5, both intrastrand and interstrand NOEs consistent with expected β -hairpin structure were observed, nevertheless NOEs consistent with the helical structure were also observed throughout the full length of a molecule. Therefore it is indicated that each of DP2, DP4, and DP5 takes both the α -helix and β -hairpin conformation concomitantly. Especially, since the spectrum of DP5 has the largest number of NOEs consistent with both the helical and β -hairpin structure among all DPs, it is evident that the helical and β -hairpin structures were most stable in DP5. To elucidate the conformation of these structures for DP5,

structural calculation was performed by using the distance restraints <1> and <3> (Conformation1: Fig.1C D, Table1) and then using <2> and <3> (Conformation2: Fig.1E F, Table1). In Conformation1, it is shown that the backbone of the residues 4-10 adopts an α -helix conformation and the side chains of Y3, W4, and H7 are clustered, being consistent with the conformation of TP. On the other hand, in Conformation2, several interacting residues are observed between strands, i.e., Y3-W16, W4-H15, and H7-Y12, showing that DP5 takes the β -hairpin structure as expected (Fig. 3F). For DP2 or DP4, each of the rmsd from the mean structure was worse than that of DP5 because the numbers of observed NOEs of DP2 or DP4 were fewer than those of DP5.

Evaluation of the structural stabilities by analyses of NOE. Population of each structured peptide was quantified using NOESY spectra in a following manner. In the initial rate approximation, the integrated cross peak (A) intensity for sufficiently slow molecular motion is expressed by

$$V_A \approx C \left(\frac{t_m}{r_A^6} \right) P_A \quad [1],$$

where C is a constant including the correlation time in a spectrum, t_m is the mixing time, r is the distance between the two interacting protons, and p is the number of pairs of the protons. Thus, if there is a reference cross peak (ref), P_A is determined by the relationship (ignoring differential internal mobility):

$$P_A = P_{ref} \left(\frac{V_A}{V_{ref}} \right) \left(\frac{r_A}{r_{ref}} \right)^6 \quad [2].$$

To satisfy the initial rate approximation, plots of the integrated cross peak versus mixing time were considered as Maclaurin series, which is expressed by

$$V = at_m + bt_m^2 + \dots \quad [3],$$

where a and b are constants. The experimentally obtained points were fitted by incorporating the first term or second term in the right side of equation [3]. The resulting first terms (at_m) were used as V_A or V_{ref} . We confirmed equation [2] by using NOEs observed between hydrogens whose distance is already known, e.g., 2(6)H and 3(5)H in Tyr3, 4H and 5H in Trp4, and methylene protons in His7. As a result of calculation ($P_A = P_{ref} = 1$) of the distance (r_A) by using integrated cross peak intensities (V_A or V_{ref}) and the distance ($r_{ref} = 2.48 \text{ \AA}$) between 2(6)H and 3(5)H in Tyr, it was shown that the calculated distances within side chains in Trp or His corresponded well with the expected values. Additionally, the helical population of TP evaluated by using this method corresponded well with that obtained by analysis of the differential scanning calorimetry (data not shown). All these results indicate quantitative validity of this method for estimation of the population of structured peptides.

By using 14 clearly separated NOE signals in group <1>, the helical populations of DPs were quantified (Table2) and compared to those of TP, in which distances related to NOE intensities were set to the average of 10 best structures of DP5 and backbone distances of 3_{10} helix²⁸. While the average helical population for DP1 (27.1 %), DP2 (24.6 %), DP3 (20.5 %), and DP4 (24.0 %) were slightly higher than or comparable to those of TP (20.5 %), it is to be noted that the population was highest for DP5 (31.7 %), which was shown to form the most stable β -hairpin structure among all peptides. The free energy of folding to the helix state ($\Delta G^0_{D \rightarrow \alpha}$) for each peptide was derived from the average of these populations (Table2).

Similarly, β -hairpin populations of DP2, DP4, and DP5 were quantified, using 4 interstrand NOEs and 7 intrastrand NOEs separately observed and the average structure

of DP5 (Table2). It is evident that β -hairpin is most stable in DP5 followed by DP4 and DP2. Incidentally in DP5, according to the calculations of the free energy of folding to the β -hairpin state ($\Delta G_{D \rightarrow \beta}^0$) of each peptide using the average of the populations (Table2), it is shown that all three states (α , β , and D) reside at similar free energy level.

Evaluation of structural stabilities by means of CD. The far-UV CD spectra of TP and DPs are displayed in Fig4A. The spectrum of individual peptide has the positive band at 229nm, indicating interactions between the aromatic chromophores ²⁹. To evaluate structural contents, we obtained the ratio $[\Theta]_{215}/[\Theta]_{198}$, which reflects the β -hairpin structure, of TP, DP1, DP2, DP3, DP4, and DP5 to be 0.18, 0.20, 0.20, 0.26, 0.33, 0.51, respectively. On the other hand, $[\Theta]_{208}/[\Theta]_{198}$, which reflects the helix structure, of TP, DP1, DP2, DP3, DP4, and DP5 became 0.44, 0.42, 0.43, 0.47, 0.54, 0.78, respectively ²⁶. Although aromatic side chains contribute to far-UV CD spectra, difference in the number of aromatic residues in each peptide was ignored, since $[\Theta]$ derived from an aromatic side chain contributes less than 5×10^3 (deg cm² dmol⁻¹) to that in the far-UV region. ³⁰. While $[\Theta]_{215}/[\Theta]_{198}$ and $[\Theta]_{208}/[\Theta]_{198}$ for DP1-3, DP4 are comparable to or slightly higher than those of TP, those for DP5 are greater than those for TP, suggesting higher content in both the α -helix and β -hairpin structure.

DP5 also showed a denaturant-induced denaturation curve while the change in CD spectrum is subtle for a mixture of TP and DP5 ¹²⁻¹⁸ (Fig.4A). The CD spectrum of the mixture, even in 6M urea, has the positive band at 229nm, which is assignable to the side chains of Y3 and W4 interacting each other, suggesting that the interaction among aromatic residues stabilizes the residual structure even in 6M urea as observed in the case of α lac:101-111 ¹⁹. On the other hand, the thermal denaturation curve of DP5 was

not notably detected (data not shown), indicating that the enthalpy of folding to the helical or β -hairpin structure is small.

Discussion

Structures and stabilities of the Designed Peptides. We added the seven residues to the C terminus of TP, which holds a fairly stable helical structure, anticipating that the peptide might be transformed into a β -hairpin structure. As a result, it was shown that the β -hairpin structure was formed in DP2, DP4, and DP5 (Fig1 E F), while the β -hairpin was missing in DP1 or DP3. The β -hairpin of DP5 is more stable than that of DP2 or DP4 judging from the greater number of NOEs in group <2> (Fig 3 B), free energies derived from NOE analyses (Table2), and the ellipticity obtained from CD measurements. In DP5, the stabilization of the β -hairpin appears to be maintained by hydrophobic interactions including side chains of aromatic residues, Y3, W4, H7, Y12, H15, and W16 (Fig. 3F).

On the other hand, NOEs corresponding to fully helical structure, which originally existed in TP, was also shown to be existed in DP2, DP4, and DP5 (Fig.3, Table2). In fact, the helices in DP2, DP4 and DP5 were more stable than that of TP, among which DP5 was the most stable. These experimental results indicate that when an amino acid sequence is added to the C terminus of a helical peptide, stabilization of newly formed secondary structure does not result in destabilization of the original secondary structure. The stabilization of the helix is realized by not only helix formation in the added sequence but also by further stabilization in the pre-existing helical portion in residues 1-7 (Fig. 3A).

Change in stability caused by interaction between fragments. To clarify the underlying principle that control the stability caused by addition of residues, the sequence of DPs is hypothetically divided into two portions; Fragment1 and Fragment2 (Fig.4B). In DP2, DP4 and DP5, combining Fragment1 with Fragment2 resulted in formation of β -hairpin manifested by increment of intrastrand NOEs in Fragment1 with concomitant appearance of interstrand NOEs (Fig3.C D). Therefore, the stabilization of the β -hairpin structures of DP2, DP4, and DP5 indicates that the β -strand in Fragment1 in each DP is more stabilized than that in TP because Fragment2 added to the C terminus of Fragment1 interacts with Fragment1.

For helix formation, we also divide $\Delta G_{D \rightarrow \alpha}^0$ of a DP, which ranges from 0.04 to 0.81 kcal/mol, into the free energy of folding to the helical state of residues 1-11 in isolation ($\Delta G_{D \rightarrow \alpha, a}^0$), that of residues 12-18 ($\Delta G_{D \rightarrow \alpha, b}^0$) in isolation, and the free energy caused by interaction between residues 1-11 and 12-18 ($\Delta G_{D \rightarrow \alpha, c}^0$). It is natural to set $\Delta G_{D \rightarrow \alpha, a}^0$ of each DP to 0.79 (kcal/mol), which is the free energy of folding to the helical state of TP (Table2), because the sequence of Fragment1 of all DPs is identical to that of TP. In addition, we can assume that $\Delta G_{D \rightarrow \alpha, b}^0$ should be larger than a few kcal/mol, since it is confirmed that even DP5¹²⁻¹⁸, of which helical structure is the most stable in all DPs, is unfolded according to NOESY and ROESY measurements (data not shown). Taken together it is shown that $\Delta G_{D \rightarrow \alpha, c}^0$ of DP2, DP4, and DP5 should be largely negative and contribute to the stabilization of the helical structure, although we first expected destabilization of the helix and stabilization of the β -hairpin. In fact, among DP2, DP4, and DP5, the more stabilized the β -hairpin structure becomes, the more stabilized the helical structure is (Table2). It is thus indicated that when one fragment is added to the

C terminus of another fragment, the interaction free energy, which arises from combining the two peptide fragments, does not always contribute to the structural stability in favor of stabilization of only one (in this case β -strand) structural conformation.

The increased stabilizations of the helical states of DP2, DP4, and DP5 can be accounted for by studies about alanine-based peptides³¹⁻³⁴. It is suggested that while a peptide that has a sequence of seven sequential alanine residues exists in an ensemble of unfolded conformations including polyproline II conformation^{31,34}, α helix becomes dominant for a peptide of 13 sequential alanine residues³². These experimental results indicate that when a polypeptide sequence is elongated, the backbone prefers α helical structure to the unfolded state. Thus, increased stabilities of the helical states of DP2, DP4, and DP5 are possibly attributed to this helical preference, which resulted from adding the seven residues to the C terminus of TP. Initially, we expected that poor stability (0.79 ± 0.21 kcal/mol) of the helical structure of TP would be overwhelmed by the strong stabilization of β -hairpin structure through favorable interactions such as π - π between Fragment1 and Fragment2. Nevertheless, the experimental results show that while the expected β -hairpin structure is stabilized, the helical structure in TP also stabilized in DP2, DP4, and DP5. It is thus concluded that in order to design a peptide that takes only a stable β -hairpin structure, we need a sequence whose fragments do not have the helical structure when isolated.

On the rule of determining and designing a stable secondary structure. The intrinsic rules of determining secondary structures of natural proteins are unknown, partly because the conformation of a protein fragment in isolation does not generally

form the same secondary structure observed in the native protein but usually unfolded. In this study, the interaction term $\Delta G_{D \rightarrow \alpha, c}^0$ did not contribute to destabilization of the original helical structure while the β -hairpin structure was indeed stabilized by the interaction between Fragment1 and Fragment2. This fact indicates that when a protein sequence is divided into several peptide fragments, a secondary structural unit that is stabilized in an independent peptide fragment is not always destroyed by newly induced interaction between the fragments. Therefore, we propose a presumption that one fragment that has only a secondary structure stabilized by the tertiary interaction with another fragment, as observed in natural proteins, has to be unfolded in isolation. We can apply this rule in not only designing artificial proteins but also analyzing folding of natural proteins.

The finding of the ‘chameleon’ sequence coincides with our experimental results in that a peptide has potential ability to have various conformations³⁵, when in the context of different environments. In addition, the presumption is supported by the fact that the ‘chameleon’ sequence, which has the α helix or β strand formed in a context-dependent manner, is unfolded in isolation. Similarly, two mutants (G_Y , G_{Hel}) of GB1 domain whose α -helix is replaced by a sequence that corresponds to a β -hairpin that is stable in isolation, fold into native-like structures, although T_m values of G_Y , G_{Hel} (44.4, 24.4°C, respectively) are drastically smaller than that of GB1 (78.9°C)³⁶. It is thus indicated that the nonnative β -hairpin structures, which are stable in isolation, caused destabilization of the native structure of GB1. In other example, while β -lactoglobulin is a predominantly β -sheet protein despite its amino acid sequence having high theoretical helicity³⁷⁻⁴⁰, its peptide fragments that form the β -strand in the protein and have high helicity are disordered in isolation in aqueous solution. This result would also support

our presumption in that even fragments that have high potential ability to have helical structure is unstructured in isolation.

The architectural principle of protein structure is very complex and physicochemical factors contributing to protein stabilization cannot, thus, be represented as linear combination. Even when the propensities of individual secondary structures have been revealed, the intrinsic rules of determining tertiary interactions between secondary structures in a protein are still unknown. Here we proposed one approach of understanding these nonlinear interactions by dividing or adding fragments to the already existed secondary structure. This approach also produced quite an intriguing peptide molecule that forms equally populated helix and sheet, which can only be realized in designed proteins.

Acknowledgements. This work was supported in part by grants from JST (to AT).

References

1. Wojcik J, Altman KH, Scheraga HA. Helix-coil stability constants for the naturally occurring amino acids in water. XXIII. Proline parameters from random poly (hydroxybutylglutamine-co-L-proline). *Bipolymers* 1990;30:121-34.
2. Lyu PC, Liff MI, Marky LA, Kallenbach NR. Side chain contributions to the stability of alpha-helical structure in peptides. *Science* 1990;250(4981):669-73.
3. O'Neil KT, DeGrado WF. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* 1990;250(4981):646-51.
4. Padmanabhan S, Marqusee S, Ridgeway T, Laue TM, Baldwin RL. Relative helix-forming tendencies of nonpolar amino acids. *Nature* 1990;344(6263):268-70.
5. Horovitz A, Matthews JM, Fersht AR. Alpha-helix stability in proteins. II. Factors that influence stability at an internal position. *J Mol Biol* 1992;227(2):560-8.
6. Blaber M, Zhang XJ, Matthews BW. Structural basis of amino acid alpha helix propensity. *Science* 1993;260(5114):1637-40.
7. Chou PY, Fasman GD. Conformational parameters for amino acids in helical,

- beta-sheet, and random coil regions calculated from proteins. *Biochemistry* 1974;13:211-22.
8. Minor DL, Jr., Kim PS. Measurement of the beta-sheet-forming propensities of amino acids. *Nature* 1994;367(6464):660-3.
 9. Kim CA, Berg JM. Thermodynamic beta-sheet propensities measured using a zinc-finger host peptide. *Nature* 1993;362(6417):267-70.
 10. Minor DL, Jr., Kim PS. Context is a major determinant of beta-sheet propensity. *Nature* 1994;371(6494):264-7.
 11. Dalal S, Balasubramanian S, Regan L. Protein alchemy: changing beta-sheet into alpha-helix. *Nat Struct Biol* 1997;4(7):548-52.
 12. Gill SC, von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* 1989;182(2):319-26.
 13. Piotto M, Saudek V, Sklenar V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* 1992;2(6):661-5.
 14. Wuthrich K. *NMR of Proteins and Nucleic Acids*. Wiley, New York; 1986.
 15. Dingley AJ, Mackay JP, Chapman BE, Morris MB, Kuchel PW, Hambly BD, King GF. Measuring protein self-association using pulsed-field-gradient NMR spectroscopy: application to myosin light chain 2. *J Biomol NMR* 1995;6(3):321-8.
 16. Stejskal EO, Tanner JE. Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *J. Chem. Phys.* 1965;42:288-92.
 17. Wilkins DK, Grimshaw SB, Receveur V, Dobson CM, Jones JA, Smith LJ. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* 1999;38(50):16424-31.
 18. Guntert P, Mumenthaler C, Wuthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* 1997;273(1):283-98.
 19. Demarest SJ, Fairman R, Raleigh DP. Peptide models of local and long-range interactions in the molten globule state of human alpha-lactalbumin. *J Mol Biol* 1998;283(1):279-91.
 20. Demarest SJ, Hua Y, Raleigh DP. Local interactions drive the formation of nonnative structure in the denatured state of human alpha-lactalbumin: a high resolution structural characterization of a peptide model in aqueous solution. *Biochemistry* 1999;38(22):7380-7.
 21. Simpson ER, Meldrum JK, Searle MS. Engineering diverse changes in beta-turn propensities in the N-terminal beta-hairpin of ubiquitin reveals significant effects on stability and kinetics but a robust folding transition state. *Biochemistry* 2006;45(13):4220-30.

22. Du D, Tucker MJ, Gai F. Understanding the mechanism of beta-hairpin folding via phi-value analysis. *Biochemistry* 2006;45(8):2668-78.
23. Alba Ed, Jimenez MA, Rico M. Turn Residue Sequence Determines Beta-Hairpin Conformation in Designed Peptides. *J. Am. Chem. Soc.* 1997;119:175-183.
24. Wouters MA, Curmi PM. An analysis of side chain interactions and pair correlations within antiparallel beta-sheets: the differences between backbone hydrogen-bonded and non-hydrogen-bonded residue pairs. *Proteins* 1995;22(2):119-31.
25. Cochran AG, Skelton NJ, Starovasnik MA. Tryptophan zippers: stable, monomeric beta-hairpins. *Proc Natl Acad Sci U S A* 2001;98(10):5578-83.
26. Pastor MT, Lopez de la Paz M, Lacroix E, Serrano L, Perez-Paya E. Combinatorial approaches: a new tool to search for highly structured beta-hairpin peptides. *Proc Natl Acad Sci U S A* 2002;99(2):614-9.
27. Griffiths-Jones SR, Searle MS. Structure, Folding, and Energetics of Cooperative Interactions between the beta-Strands of a de Novo Designed Three-Stranded Antiparallel beta-Sheet Peptide. *J. Am. Chem. Soc.* 2000;122:8350-56.
28. Wuthrich K, Billeter M, Braun W. Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton-proton distances. *J Mol Biol* 1984;180(3):715-40.
29. Grishina IB, Woody RW. Contributions of tryptophan side chains to the circular dichroism of globular proteins: exciton couplets and coupled oscillators. *Faraday Discuss* 1994(99):245-62.
30. Chakrabartty A, Kortemme T, Padmanabhan S, Baldwin RL. Aromatic side-chain contribution to far-ultraviolet circular dichroism of helical peptides and its effect on measurement of helix propensities. *Biochemistry* 1993;32(21):5560-5.
31. Shi Z, Olson CA, Rose GD, Baldwin RL, Kallenbach NR. Polyproline II structure in a sequence of seven alanine residues. *Proc Natl Acad Sci U S A* 2002;99(14):9190-5.
32. Spek EJ, Olson CA, Shi Z, Kallenbach NR. Alanine Is an Intrinsic alpha-Helix Stabilizing Amino Acid. *J. Am. Chem. Soc.* 1999;121:5571-72.
33. Baldwin RL. A new perspective on unfolded proteins. *Adv Protein Chem* 2002;62:361-7.
34. Makowska J, Rodziewicz-Motowidlo S, Baginska K, Vila JA, Liwo A, Chmurzynski L, Scheraga HA. Polyproline II conformation is one of many local conformational states and is not an overall conformation of unfolded peptides and proteins. *Proc Natl Acad Sci U S A* 2006;103(6):1744-9.
35. Minor DL, Jr., Kim PS. Context-dependent secondary structure formation of a designed protein sequence. *Nature* 1996;380(6576):730-4.

36. Cregut D, Civera C, Macias MJ, Wallon G, Serrano L. A tale of two secondary structure elements: when a beta-hairpin becomes an alpha-helix. *J Mol Biol* 1999;292(2):389-401.
37. Papiz MZ, Sawyer L, Eliopoulos EE, North AC, Findlay JB, Sivaprasadarao R, Jones TA, Newcomer ME, Kraulis PJ. The structure of beta-lactoglobulin and its similarity to plasma retinol-binding protein. *Nature* 1986;324(6095):383-5.
38. Kuwajima K, Yamaya H, Miwa S, Sugai S, Nagamura T. Rapid formation of secondary structure framework in protein folding studied by stopped-flow circular dichroism. *FEBS Lett* 1987;221(1):115-8.
39. Shiraki K, Nishikawa K, Goto Y. Trifluoroethanol-induced stabilization of the alpha-helical structure of beta-lactoglobulin: implication for non-hierarchical protein folding. *J Mol Biol* 1995;245(2):180-94.
40. Hamada D, Segawa S, Goto Y. Non-native alpha-helical intermediate in the refolding of beta-lactoglobulin, a predominantly beta-sheet protein. *Nat Struct Biol* 1996;3(10):868-73.
41. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. *J. Appl. Cryst.* 1993;26:283-91.
42. Tanford C. Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv Protein Chem* 1970;24:1-95.
43. Jackson SE. How do small single-domain proteins fold? *Fold Des* 1998;3(4):R81-91.

Table 1. NMR structural statistics for Target Peptide and Design Peptides

Parameter	Target Peptide	DP5 (Conformation1)	DP5 (Conformation2)
Rmsd of from the mean structure (Å)			
Backbone atoms	0.23 ± 0.12	0.30 ± 0.12	0.83 ± 0.23
All heavy atoms	0.85 ± 0.22	0.77 ± 0.24	1.33 ± 0.28
Ramachandran analysis (%)			
Most favored regions	75.0	60.7	18.7
Additional allowed regions	25.0	38.0	62.7
(residues used for rmsd calculation)	(3-9)	(3-16)	(3-16)

Ramachandran analysis was evaluated by using the program PROCHECK ⁴¹.

Table2. Structural population (%) and $\Delta G_{D \rightarrow \alpha (\beta)}$ (kcal/mol) of Target Peptide and Design Peptides

	α Population (%)	$\Delta G_{D \rightarrow \alpha}^0$ (kcal/mol)	β Population (%)	$\Delta G_{D \rightarrow \beta}^0$ (kcal/mol)
TP	20.5 ± 6.4	0.79 ± 0.21	0	-
DP1	27.1 ± 6.4	0.57 ± 0.16	0	-
DP2	24.6 ± 7.9	0.49 ± 0.25	19.4 ± 9.8	0.68 ± 0.38
DP3	20.5 ± 8.5	0.81 ± 0.28	0	-
DP4	24.0 ± 7.9	0.38 ± 0.30	30.9 ± 14.1	0.27 ± 0.38
DP5	31.7 ± 12.1	0.04 ± 0.42	34.9 ± 17.8	0.03 ± 0.51

$\Delta G_{D \rightarrow \alpha}^0$, $\Delta G_{D \rightarrow \beta}^0$ is the free energy of folding to halical, β -hairpin state, respectively.

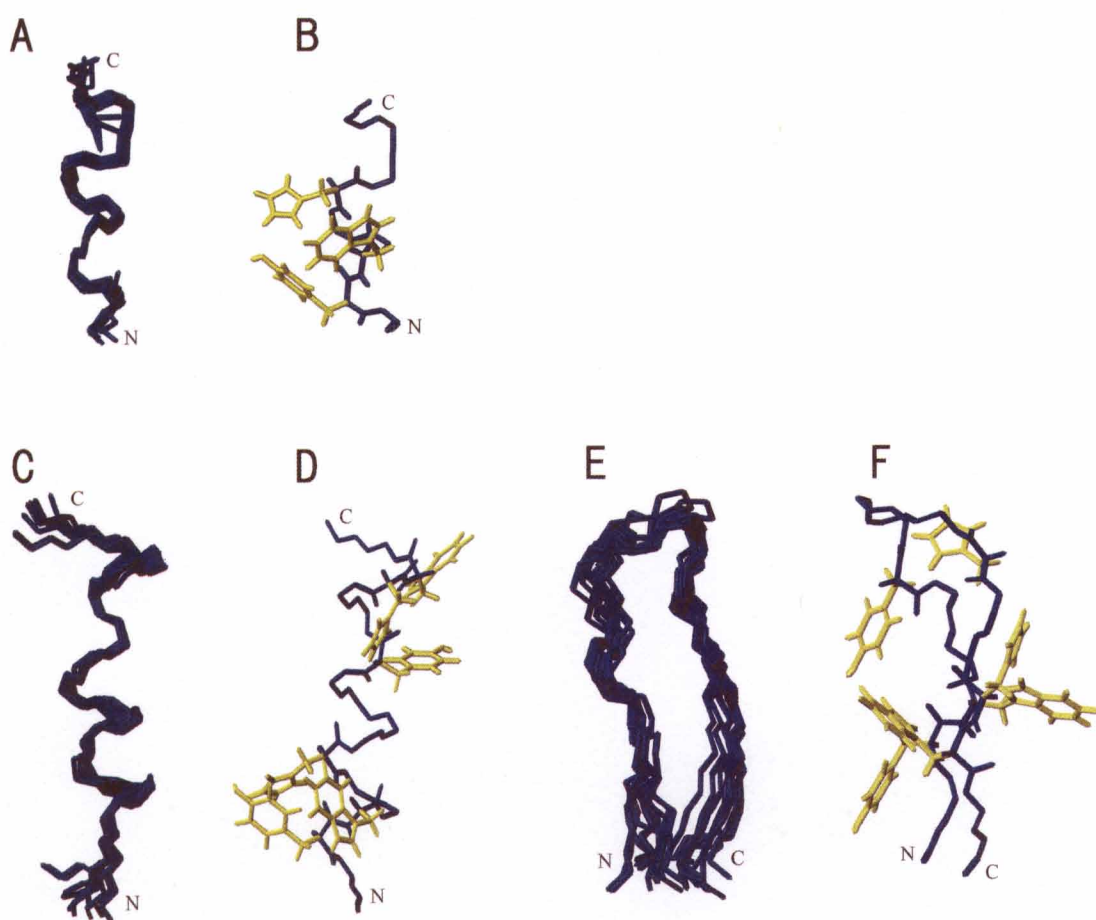


Fig.1 NMR structures of TP (A, B) and DP5 (C, D, E, F). (A) Backbone traces of the 10 best structures. (B) Minimized mean of the 10 best structures shown in (A). (C) Backbone traces of the 10 best structures that were calculated by using the distance restraints <1> and <3> (Conformation1, see text for definition of restraints). (D) Minimized mean of the 10 best structures shown in (C). (E) Backbone structures of the 10 best structures that were calculated by using the distance restraints <2> and <3> (Conformation2). (F) Minimized mean of the 10 best structures shown in (E). Residues drawn in yellow are side chains of Y3, W4, H7, Y12, H15, and W16 in TP or DP5.

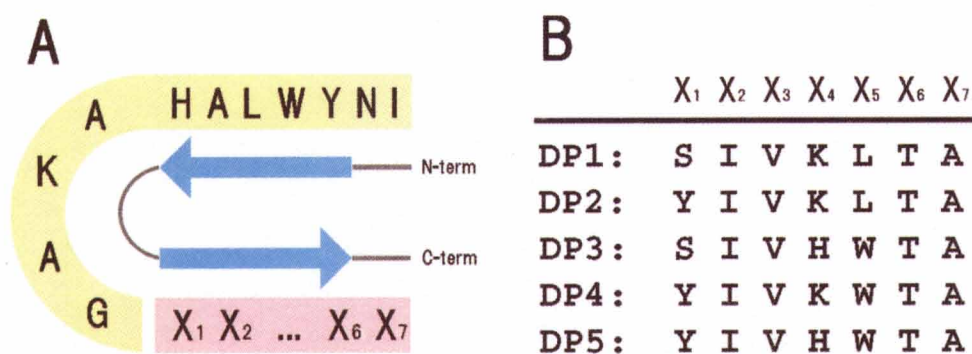


Fig.2 (A) Schematic diagram of the design of a β -hairpin structure. Target Peptide region is shaded in yellow while extra region (X₁, X₂, ... X₆, X₇) is shaded in magenta. (B) Sequences of Design Peptides (DP1-5)

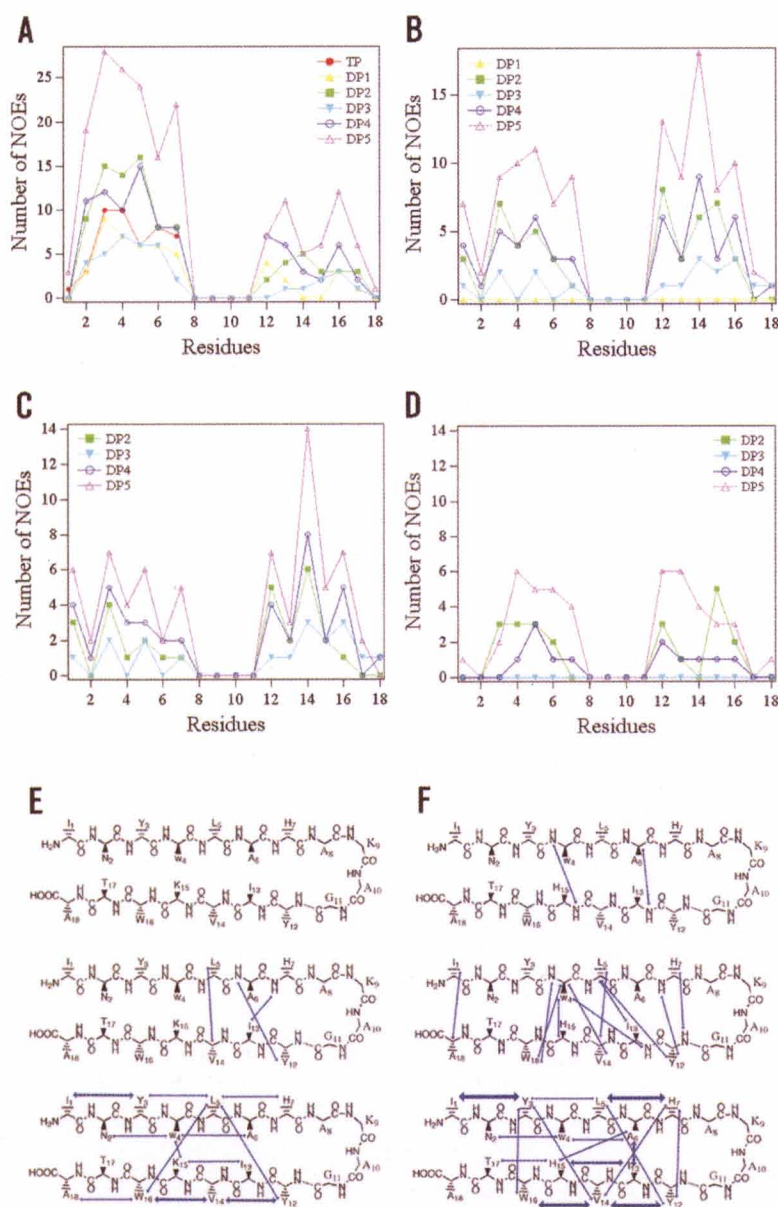


Fig.3 (A) (B) (C) (D): plots of the number of distributed NOEs per residue, (E) (F): summary of long and medium range NOEs consistent with a β -hairpin structure. (A) NOEs consistent with helical structure only: NOEs of $d_{NN}(i, i+1)$, $d_{NN}(i, i+2)$, $d_{\alpha N}(i, i+2)$, and those observed between residues i and $i+3$, between residues i and $i+4$, and between side chains of residues i and $i+1$. (B) NOEs consistent with a β -hairpin structure only: interstrand ones consistent with expected β -hairpin conformation, intrastrand ones observed between side chains of residues i and $i+2$. (C) intrastrand NOEs in those consistent with a β -hairpin structure. (D) interstrand NOEs in those consistent with a β -hairpin structure. (E) and (F): NOEs observed in DP4 and DP5, respectively, superimposed on the expected hairpin structure. Upper: backbone-backbone NOEs, middle: backbone- side chain NOEs, bottom: side chain-side chain NOEs. Thin, middle, and thick arrow indicate that the number of NOEs is 1-2, 3-4, and 5-7, respectively.

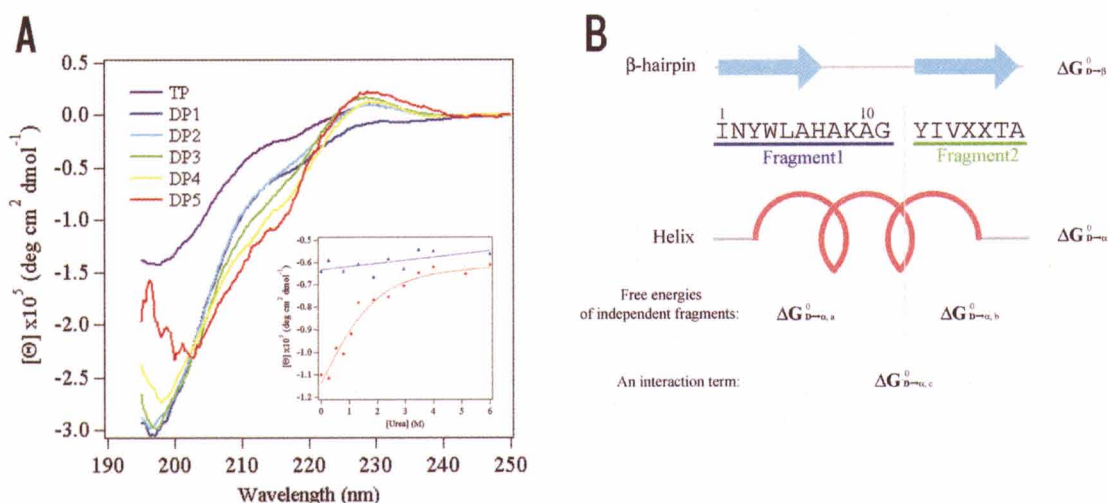


Fig.4 (A) Far-UV CD spectra of Target Peptide (TP) and Designed Peptides (DP1-5). (Inset) Urea-induced denaturation of 130μM DP5 (● in red), 130μM Target Peptide + 130μM residues 12-18 of DP5 (▲ in blue) by the change in the CD signal at 215nm. The fitted curve was drawn in the following manner. By assuming two-state transition between β-hairpin and denatured state, the denaturation curve can be fitted according to

the following linear relationship: $\Delta G_{D \rightarrow \beta}^0 = \Delta G_{D \rightarrow \beta}^{0, H_2O} + m_{eq}[\text{urea}]$, where $\Delta G_{D \rightarrow \beta}^{0, H_2O}$, which

was independently obtained by means of NMR analysis to be 0.03kcal/mol (Table 2), is the free energy of folding in the absence of denaturant and m_{eq} is a constant of proportionality for the dependence of the free energy change on denaturant concentration⁴². When the pretransition baseline (θ_N) was assumed to be constant (i.e., $\theta_N = A$, where A is a fitting variable) and the posttransition baseline (θ_D) was treated as a linear function of temperature (i.e., $\theta_D = B + CT$, where B and C are predetermined according to the least-squares fitting), m_{eq} obtained from the fitting is 0.57 ± 0.06 (kcal/mol), which is smaller than m_{eq} values determined for natural proteins⁴³.

(B) Schematic diagram of division of the free energy of folding to the helix state of each

Design Peptide ($\Delta G_{D \rightarrow \alpha}^0$). When the sequence of a Design Peptide is divided into

residues 1-11 (Fragment1) and residues 12-18 (Fragment2), the free energy of independent Fragment1, Fragment2, the interaction energy between Fragment1 and

Fragment2 is defined as $\Delta G_{D \rightarrow \alpha, a}^0$, $\Delta G_{D \rightarrow \alpha, b}^0$, $\Delta G_{D \rightarrow \alpha, c}^0$, respectively. Thus $\Delta G_{D \rightarrow \alpha}^0$

is expressed by $\Delta G_{D \rightarrow \alpha}^0 = \Delta G_{D \rightarrow \alpha, a}^0 + \Delta G_{D \rightarrow \alpha, b}^0 + \Delta G_{D \rightarrow \alpha, c}^0$.

Chapter 2.

Protein segment that drastically decreases the solubility induces formation of the whole protein structure

Abbreviations: IP: Initial Protein, FP: Final Protein, NOE: nuclear overhauser effect, NMR: nuclear magnetic resonance, CD: circular dichroism.

Abstract

Recent protein studies suggest that each of well-packed structures of natural proteins and computationally created proteins is stabilized by a number of intra or intermolecular interactions compensating marginally for the entropy cost, thus the number of proteins having well-packing structures is extremely small compared to that of every possible primary sequence. Nevertheless, natural proteins have evolved to well-packed structures, presumably because that proteins having “minimal structures” necessary to exercise biological functions minimally are widely distributed in the possible primary sequence space. This presumption means that an unfolded protein can have a minimal structure by minor change in the primary sequence, and then the amino acid change could be a minimum principle to form protein structure. We thus tried to transform a fully unfolded protein consisting of 25 residues, defined as “Initial Protein” or “IP”, into a fully structured protein, defined as “Final Protein1” or “FP1”, by adding a peptide fragment to the C-terminus of the original protein. For choice of amino acids of the peptide fragment, we deduced that adding a peptide fragment to decrease the protein solubility results in induction of formation of a minimal structure. This is because protein folding is a phenomenon in which a protein excludes a part of the molecule from solvent in a geometry-specific manner, and then, without any regard for protein structural specificity, a state in which a protein molecule is the most hidden from solvent is the precipitating state. As an experimental result, a whole protein structure including intermolecular interactions, defined as “soft-packed structure”, was formed with a drastic decrease in the protein solubility from $> 2.2\text{mM}$ to $\sim 10\mu\text{M}$. Additionally, hydrophilic replacements of the added peptide fragment that yield an increase in the solubility compared to FP1

resulted in drastic destabilization of the protein structure. The fact that the short protein segment to make the whole protein insoluble conferred the ability to form the soft-packed structure to the remaining long segment verified that all amino acid residues composing a protein need not to be correctly chosen to form soft-packed structures. Furthermore, soluble IP easily became insoluble by addition of the fragment equivalent to $\sim 1/4$ length of IP, suggesting that soft-packed structures observed in FP1 can be formed much more frequently.

Introduction

To elucidate architectural principle of protein structure, the relationship between protein sequence and the tertiary structure has been studied from various perspectives. Previously, determining factors related to protein sequence of the structural stabilizing mechanism have been suggested mostly by studies of stability and folding kinetics of natural proteins and the mutagenesis studies ¹⁻³. Recently, computational protein designs have advanced to provide new insights into the determinants of protein structure, stability, and folding ⁴. Computational methods for identifying amino acid sequences compatible with a known target structure have allowed redesign of naturally occurring proteins ⁵⁻⁷. Therefore, finding novel sequences that have a known protein structure has become possible in some cases ^{6,8}. In addition, mutating partially buried polar residues of wild-type proteins to hydrophobic residues by the redesign methods has shown that an increase in hydrophobic surface area burial has resulted in an increase in the protein stability ⁹⁻¹¹. On the other hand, proteins with novel structures have been created by

methods of computational design^{12,13}, indicating that a large number of protein folds have not been sampled by nature, in evolutionary history. These successful computational designs to create novel protein sequences and structures not only suggest that the potential function guiding the design process captures much of the important physical chemistry, but make it possible to elucidate the properties of natural proteins, which have been selected by various evolution pressures, by comparing the properties of artificially designed counterparts.

Above-described studies suggest that each of well-packed structures of natural proteins and computationally created proteins is stabilized by a number of intra or intermolecular interactions compensating marginally for the entropy cost, thus, the number of proteins having well-packed structures seems to be extremely small compared to that of every possible primary sequence. How has nature created proteins having well-packed structures in case sequences that fold into well-packed structures are extremely rare? If most of every possible primary sequence are unfolded, natural proteins should not have evolved to well-packed structures, presuming that proteins having “minimal structures” necessary to exercise biological functions minimally should be widely distributed in the possible primary sequence space. This presumption means that an unfolded protein can have a minimal structure by minor change in the primary sequence, and then the amino acid change could be a minimum principle to form protein structure. Thus, we experimentally transform a fully unfolded protein (defined as “Initial Protein” or “IP”) into a fully structural protein (defined as “Final Protein” or “FP”) by adding a peptide fragment to the C terminus of the original protein. The peptide fragment should have an ability to induce formation of the whole protein structure, thus, amino acids of the peptide fragment need to be chosen prudently. Since

protein folding is a phenomenon in which a protein excludes a part of the molecule from solvent in a geometry-specific manner, driving force of protein folding could be regarded as hydrophobicity, without any regard for protein structural specificity¹⁴. Thus, it is naively indicated that the more insoluble a protein becomes, the more stable minimal structures become since the most hidden state from solvent is the precipitating state. Therefore we deduce that induction of formation of minimal structures needs to add a peptide fragment that decreases the protein solubility. We thus examined the solubility dependence of the structural stability, and then we discussed necessary hydrophobicity of the added peptide fragment to induce forming of the whole protein structure. The finding of a soft-packed structure that involves intermolecular interactions, observed as minimal structure, not only suggests a factor necessary to induce a whole protein fold, but leads to a minimal principle to form the protein structure.

Materials and Methods

Protein synthesis and purification. Proteins were synthesized by Pioneer Peptide Synthesis System (PerSeptive Biosystems, CA, USA) using Fmoc solid-phase chemistry and were cleaved from the resin with a solution containing 82.5% (vol/vol) trifluoroacetic acid (TFA), 5% H₂O, 5% thioanisole, 2.5% 1,2-ethanedithiol, and 0.8M phenol. Individual proteins were purified by reverse-phase HPLC (acetonitrile/H₂O/0.1%TFA). Peptide identity was confirmed by laser desorption time of flight mass spectrometry, AXIMA-CFR (SHIMADZU, Kyoto, Japan). Protein samples for all studies were lyophilized and stored in the anaerobic condition.

Following experiments of all were performed under a N₂ atmosphere with the use of buffers deoxygenated with N₂ to prevent cysteine oxidation.

Circular Dichroism (CD) measurements. Spectra were acquired at 20°C on a Jasco J-720 CD spectropolarimeter with 0.1, 0.2, 1, 5, 10mm pathlength cuvettes on protein samples of 0.004-3mM concentration. After each protein was dissolved in a buffer containing 25mM acetic acid, 2-4mM NaOH, and 50mM NaCl in 90% H₂O and 10% D₂O, the solution was adjusted to pH 3.0 ± 0.1 with NaOH or HCl. Protein concentrations were determined by measurements of protein sulphydryls with Ellman's reagent as described ¹⁵.

Ultracentrifuge measurements. Each protein sample was prepared as described in materials and methods of CD measurements. Sedimentation velocity and sedimentation equilibrium measurements were performed using a Beckman-Coulter Optima XL-1 analytical ultracentrifuge (Fullerton, CA) with an An-60 rotor and two-channel

charcoal-filled Epon cells at 20°C and pH3.0 $\pm$ 0.1. Sedimentation equilibrium was measured at 0.9mM, 1.5mM, and 3.0mM FP1 concentrations while sedimentation velocity was measured at 2.3mM and 3.0mM FP1 concentrations. The data were analyzed using the software Ultrascan 6.01 (www.ultrascan.uthscsa.edu/).

NMR spectroscopy. NMR spectra were performed on a Bruker DMX-750 spectrometer at 20°C on protein samples of 0.9-3.0mM concentration. After each protein was dissolved in a buffer containing 25mM acetic acid, 0-4mM NaOH, and 50mM NaCl in 90% H₂O and 10% D₂O, the solution was adjusted to objective pH with NaOH or HCl. Pulsed-field gradient NMR spectra were acquired at 20°C and pH3.0 on 0.5mM FP1 concentration, at which CD and sedimentation equilibrium measurements suggested that FP1 was the monomer state, in a acetate buffer containing 40mM 1,4-dioxane in 90% H₂O and 10% D₂O. All chemical shifts were referenced to the sodium salt of trimethylsilylpropionate (TSP). Pulsed-field gradient NMR spectroscopy, double quantum filtered correlation spectroscopy (2QF COSY), total correlation spectroscopy (mixing time 80ms), and nuclear Overhauser effect (NOE) spectroscopy (mixing time = 200ms) experiments were performed and water suppression was achieved by selective presaturation or field-gradient pulses ¹⁶. The proton resonances were assigned by the sequential assignment procedure ¹⁷. Populations of structured molecules were obtained by analyzing NOESY spectra at 3mM protein concentration and mixing time of 100, 150, 200, 250, and 300ms.

Structure calculations. Distance restraints were obtained by converting integrated NOE peak intensities into distance upper limits, using the macro CALIBA in DYANA¹⁸.

Standard pseudo atom distances were used when they were needed. Torsion angle constraints for ϕ were determined from $^3J_{N\alpha}$. They were then classified into three categories: $-120^\circ \pm 70^\circ$, $-120^\circ \pm 50^\circ$, and $-120^\circ \pm 40^\circ$ corresponding to $^3J_{N\alpha} < 7.5$, 7.5-8.5, and >8.5 Hz, respectively. With a cutoff of 0.2Å for upper bound NOE violations, a total of 50 structures was generated by using DYANA and the 10 lowest energy structures were selected to represent three-dimensional structures.

Solubility measurements. Solubility experiments were performed using saturated protein solutions. Samples of 200-400μL protein suspensions in buffers containing 25mM phosphate and 50mM NaCl in 90% H₂O and 10% D₂O were mixed by pipetting, then incubated for twenty minutes, by taking account of cysteine oxidation, at 25°C. After centrifugation, the pH and the concentration of the individual supernatant solutions were measured.

Analysis of the protein solubility. The chemical potential of a solute p in a real solution ($\mu_{p(sol)}$) is generally expressed by

$$\mu_{p(sol)} = \mu_{p(sol)}^0 + RT \ln \gamma_p S_p \quad [1],$$

where $\mu_{p(sol)}^0$ is the chemical potential in the ideal solution at a standard concentration of p , R is gas constant, T is absolute temperature, γ_p is the activity coefficient of p , and S_p is the concentration of p . As a first approximation, if p is present as p^{z+} ion impenetrable to the solvent, for compact protein ions, $\mu_{p(sol)}^0$ could be divided into the free energy of solvation (ΔG_{solv}^0) that depends on the valence of the ion, Z , and a term independent on the charge of the ion ($\mu_{p(sol)}^{\prime 0}$):

$$\mu_{p(sol)}^0 = \mu_{p(sol)}^{0'} + \Delta G_{solv,p}^0 \quad [2].$$

ΔG_{solv}^0 in equation [2] is expressed by Born equation:

$$\Delta G_{solv,p}^0 = -\frac{Z^2 e^2 N_a}{8\pi\epsilon_0 r_p} \left(1 - \frac{1}{\epsilon_r}\right) \quad [3],$$

where e is charge of an electron, N_a is Avogadro's number, ϵ_0 is electric constant, ϵ_r is relative permittivity, and r_p is the radius of the ion. In addition, γ_p in equation [1] is expressed by extended Debye-Hückel law:

$$\log \gamma_p = -\frac{AZ^2 \sqrt{I}}{1 + Br_p \sqrt{I}} \quad [4],$$

where A and B are constants, and I is the ionic strength in the solution. In saturated solution, since $\mu_{p(sol)}$ is equal to the chemical potential of p in the solid ($\mu_{p(s)}$), we can get for an equation of the solubility of p by using equations [1]-[4]:

$$\ln S_p = \left(\frac{\ln 10 A \sqrt{I}}{1 + Br_p \sqrt{I}} + \frac{C}{r_p} \right) Z^2 + \frac{\mu_{p(s)} - \mu_{p(sol)}^{0'}}{RT} \quad [5],$$

where

$$C = \frac{e^2 N_a}{8\pi\epsilon_0 RT} \left(1 - \frac{1}{\epsilon_r}\right) \quad [6].$$

Experimentally obtained plots of individual protein solubility were fitted by equation [5] by using $A=0.512$ ($l^{1/2}mol^{-1/2}$), $B=0.329$ ($\text{\AA}^{-1}l^{1/2}mol^{-1/2}$), and $C=281$ ($\text{\AA}$) at 25°C, presumed that a dissolved protein with the net charge, Z , which depended on pH of the solution, was a spherical ion with charge Z and radius r_p impenetrable to the solvent¹⁹.

Results

A sequence and structural property of Initial Protein. Among several candidates for

initial protein, we firstly chose third zinc finger domain, Sp1f3, of transcription factor Sp1, which had two histidines (His21, His25) and two cysteines (Cys5, Cys8) to bind covalently Zn^{2+} ²⁰. Sp1f3 folded into a well defined structure upon binding Zn^{2+} while was unfolded in the absence of the metal. As the sequence property of Sp1f3, the frequency of dissociative amino acid residues was especially high (Table1). Thus, to suppress the especially high hydrophilicity, the residues 26-29 (QNKK) in C terminus were removed and Lys1, Lys2, Glu7, and His17 were replaced with alanine or tyrosine. In addition, His25 was replaced with alanine to suppress interactions with small amount of metal ions in solution, resulting in the sequence of IP (Table1). In NOESY and TOCSY spectra of 3mM IP (Fig.1 A), most of NOE peaks of IP overlapped with TOCSY cross-peaks of that, indicated that most of the NOE peaks were the intraresidue NOE peaks, while a small number of non-intraresidue NOE peaks were identified as sequential $C^{\alpha}H-NH$, $C^{\beta}H-NH$, $NH-NH$ peaks, and sequential peaks related to $C^{\delta}H$ of prolines.

In addition to the NMR analyses, far-UV CD spectra of IP, which had the negative band at 200nm, did not change in the protein concentration range of 0.4-2.9mM, showing that IP was unfolded at up to 3mM (Fig.2).

A sequence and structural property of Final Protein. We attempted to add a peptide fragment to the C terminus of IP, which was unfolded up to 3mM, anticipating that the structure might transform throughout the length of the molecule. The number of additional residues was experimentally determined to six for the reason that contiguous hydrophobic residues in a protein resulted in reduction of the protein yield of chemical synthesis. By taking into account hydrophobicity of 20 common amino acids estimated

by hydration potentials of side chains of amino acids ^{21,22}, amino acids in the extra peptide fragment were chosen among Gly, Pro, Leu, Ile, Val, Ala, Phe, Cys, and Met, of which hydration potentials ($> \sim -2\text{kcal/mol}$) are notably larger than those of any other amino acids ($< \sim -5\text{kcal/mol}$). In addition, the sequence ($X_1, X_2, \dots, X_6$) of extra region was determined by that $X_1, X_2, X_3, X_4, X_4, X_5$, and X_6 in the sequence favored interactions with P, C, A, F, A, and Y in N-terminus, respectively ²³, because the interactions between N terminus and C terminus were seemed to be important to fold throughout the length of the molecule, resulting in that the sequence of the extra region in Final Protein, FP1, was FIVVAL (Table1). In a NOESY spectrum of 3mM FP1 (Fig.1 B), a number of NOE peaks including long range NOEs, *i.e.* $Y1C^\delta H-I27C^\gamma H$, $Y1C^\epsilon H-I27C^\gamma H$ (Fig.1 B), and medium range NOEs, *i.e.* cross peaks of $d_{NN}(i, i+1)$, $d_{\alpha N}(i, i+2)$, $d_{\alpha\beta}(i, i+3)$, beyond intraresidue NOEs were observed. On the other hand, in a NOESY spectrum of 0.9mM FP1, only intraresidue NOE peaks, sequential $C^\alpha H-NH$, $C^\beta H-NH$, $NH-NH$ peaks, and sequential peaks related to $C^\delta H$ of proline, of which most were also observed in the NOESY spectrum of 3.0mM IP, were observed. In a NOESY spectrum of 1.5mM FP1, long and medium range NOEs strongly observed in that of 3.0mM FP1 beyond NOEs observed in that of 0.9mM FP1 were observed.

In addition to the NMR analyses, far-UV CD spectra of FP1 showed that the shape of spectrum was dependent on the protein concentration (Fig.2 A). $[\Theta]$ value at 222nm of 0.4mM FP1 spectrum was ~ -2000 , which was close to that of 0.4mM IP, while that gained negatively to ~ -4000 with an increase in protein concentration to 3mM (Fig.2 B), showing that amount of formed secondary structure of FP1 gained as FP1 concentration increased.

Self-association property of FP1. We thus examined degree of protein association with an increase in protein concentration by measuring sedimentation equilibrium (Fig.3 A) and sedimentation velocity (Fig.3 B). At 0.9mM FP1 concentration, sedimentation equilibrium measurements showed that most of plots of the apparent molecular weight were distributed in 0-10000, in which molecular weights of a monomeric and a dimeric FP1 were 3500 and 7000, respectively, while a small number of the plots were distributed over 10000. At 1.5mM FP1 concentration, it was shown that a large number of the plots of the apparent molecular weight were distributed in 0-10000, whereas, the ratio of the plots over 10000 was higher than at 0.9mM FP1 concentration. At 3.0mM FP1 concentration, it was shown that most of plots of the apparent molecular weight were distributed in 50000-70000, in which molecular weights of an oligomer consisting of about 15-20 monomers were. The results of sedimentation velocity measurements were consistent with those of sedimentation equilibrium measurements. At 3.0mM FP1 concentration, the distribution of sedimentation coefficients showed two main peaks, one at 4.4 and the other at 4.6. These peaks could be regarded as the oligomers consisting of 15-20 monomers because it was indicated that most of FP1 molecules formed the oligomers of the apparent molecular weight of 50000-70000 by sedimentation equilibrium measurements at 3.0mM FP1 concentration. At 2.3mM FP1 concentration, the distribution mainly showed a peak at 1, which was the smallest sedimentation coefficient in all of observed sedimentation coefficients, in addition to a peak at 4.5, which could be regarded as the oligomers consisting of 15-20 monomers. A sedimentation coefficient of monomeric FP1 could be estimated by using an equation making a correlation between a sedimentation coefficient (S) and the molecular weight (M): $S = M(1-\rho v_s)D/RT$, where ρ was the density of a solution, v_s was the partial

specific volume of the solute, D was the diffusion constant of the solute, R was gas constant, and T was absolute temperature. As a results of the calculation by using $M=3500$ (g/mol), $\rho=1$ (g/cm³), $v_s=0.7$ (cm³/g), which was used as the general value of native protein ²⁴, $R=8.3$ (J/K mol), $T=293$ (K), and $D=9.3 \times 10^{-11}$ (m²/s), which was obtained by using pulsed-field gradient NMR spectroscopy as described ^{25,26}, S became 0.4, which was close to 1, indicating that the peak at 1 might be regarded as the monomeric FP1 or the oligomer close to the monomer. The results of sedimentation equilibrium and sedimentation velocity measurements indicated that the ratio of monomeric FP1 or the oligomer close to the monomer decreased with an increase in FP1 concentration, while that of the oligomers consisting of about 15-20 monomers increased.

pH dependence property of solubility and the structural stability of FP1 and the variants. The added peptide fragment to the C-terminus of IP in FP1 consisted of the hydrophobic amino acids, which had notably large hydration potentials. We thus prepared the variants of FP1 in which the hydrophobicity of the individual peptide fragment was lower than that of the peptide fragments of FP1 (named FP2 and FP3 (Table1)). The hydration potential of FP2, which was the variant Phe26Tyr of FP1, was 5.4 kcal/mol lower than that of FP1, while the hydration potential of FP3, which was the variant Phe26Tyr and Val28Ser, was 7.1 kcal/mol lower than that of FP2 ²¹. To evaluate hydrophobicity of FPs, experimentally, the pH dependence property of the solubilities was measured (Fig.4 A). The plots of the individual solubility of FPs were gradually increased with a decrease in pH in the pH range of 6.5-7.3, resulted from increment of the net charges, primarily, with protonation of an imidazole group in His21 and

associations of sulfhydryl groups in Cys5 and Cys8 (Fig.4 B). At measured all pH, the solubility of FP2 was as high as that of FP1, while the solubility of FP3 was clearly higher than the solubilities of FP1 and FP2. The plots of the solubility at pH < 6.4 were omitted because the plots at the pH range widely varied, resulted from that the slope of the solubility curve was heavy, *i.e.* the solubilities of FP1 at pH 6.3 and 3.4 were higher than 1.9mM and 11.4mM, respectively. The experimentally obtained plots of FPs were fitted by using equation [5], where r_p of FPs were set to 384Å obtained from the fitting of FP3, of which error was smallest in FPs for the reason of that the compositions of amino acids of FPs were almost identical. $(\mu_{p(s)} - \mu_{p(sol)}^{0'}) / RT$ obtained by the fitting of FP1, FP2, and FP3 were -22.8 ± 0.1 , -22.5 ± 0.1 , and -20.5 ± 0.0 , respectively.

The pH dependence property of structural stabilities of 3mM FP1 was given in Fig.5 A and B. With an increase in pH, integrated intensities of long-range NOE peaks were increased (Fig.5 A), while those of short and medium-range NOE peaks were also increased (Fig.5 B). The increment of integrated intensities of those NOE peaks with an increase in pH was also shown in FP2 and FP3 at 3mM protein concentration (data not shown).

Discussion

The solution structures of FPs. Complete assignment of the proton chemical shift resonances was achieved for FPs, excluding the amide protons of the N-termini, which were not detected due to rapid exchange with solvent. Nevertheless, some cross peaks in NOESY spectra of 3mM FPs were overlapped one another and could not be assigned for the reason that the proton chemical shifts did not disperse, as shown for proton chemical shifts of typical native proteins (Fig.1 B). A summary of torsion angle

restraints obtained from DQF-COSY spectrum of 3mM FP1 and the distance restraints obtained from clearly separated NOE peaks of 3mM FP1 is shown in Table 2. Since the sedimentation equilibrium and sedimentation velocity measurements of FP1 indicated that most of FP1 molecules formed the oligomers consisting of about 15-20 monomers at 3mM FP1 concentration, it is likely that some distance restraints were due to intermolecular interactions. The fact that intermolecular interactions of IP were not observed even at 3mM by CD measurements suggested that the intermolecular interactions were induced by added peptide fragment in FP1. On the other hand, content ratio of long-range NOEs related to the added peptide fragment to a total of long-range NOEs was 73%, which was notably higher than the content ratio of intraresidue (21%), short-range (30%), and medium-range (28%) NOEs, indicating that long-range interactions were mainly induced by the added peptide fragment. We thus deduced that long-range distance restraints were due to intermolecular interactions, while other distance restraints were due to intramolecular interactions. The final structural calculation of a FP1 molecule was performed with a total of 342 distance restraints, excluding 52 long-range distance restraints, and 25 backbone ϕ dihedral angle restraints (Table 2). The resulting r.m.s.d. from the mean structure for the backbone atoms was 2.79 ± 0.71 Å, which was worse than that of general native proteins (less than 0.5 Å) since long-range distance restraints were not included in the structural calculation. In a stereo view of the 10 best structures of FP1 (Fig6 A), it was shown that the backbone of residues Phe12-Ile22 adopted α -helix. On the other hand, the backbone of residues Asp16-Gln26 of Sp1f3-Zn²⁺ complex formed α -helix or 3_{10} -helix while that of residues Phe12-Ser15 formed the turn between the second β -strand and the helix, indicating that the fragment added to the C-terminus of IP in FP1 induced the α -helix formation of

portion of FP1 that had potential ability to form α -helix. Excluded long-range distance restraints were represented by using three of the lowest energy structures of FP1 (Fig.6 B). The long-range interactions of FP1 could be divided into three categories: (1) interactions between N-terminus region (Tyr1-Cys8) and C-terminus region (His21-Ala30), (2) interactions between the N-terminus region and middle region (Arg11-Arg14), (3) interactions between the middle region and the C-terminus region. The N-terminus region and the C-terminus region mainly consisted of hydrophobic amino acids, and the middle region also included hydrophobic amino acids of Phe12 and Met13. Therefore, the intermolecular interactions consisted mainly of hydrophobic interactions. Amount of the formed secondary structure gained with an increase in FP1 concentration by CD analyses, indicating that the hydrophobic interactions between FP1 molecules stabilized the local structure including α -helix of a FP1 molecule.

The most appropriate physicochemical factor that determines structural stabilities of FPs. Populations of structured molecules at each pH of 3mM FPs were quantified by using 6 short or medium-range NOE peaks clearly separated and distances of 10 best structures of FP1²⁷. The average of the populations at each pH was shown in Fig.5 C. The populations of FP1 and FP2, which were close to 0 at pH 2.5, increased to about 0.4 with an increase in pH to 4.2, whereas the structural stabilities were not measured at pH > 4.2 because of an increase in the aggregation rate. On the other hand, the population of FP3 was close to 0 at pH 3.8 while that increased to about 0.4 with an increase in pH to 5.6. Since the protein solubility, the hydration potential, and the structural stability of FPs were largely dependent on pH, we examined the protein solubility or the hydration potential dependence property of the structural stability. On this occasion, the structural

stability and the protein solubility were transformed into the folding free energy (ΔG_f) and the dissolution free energy (ΔG_{dissol}), respectively. Firstly, using the average of the populations, the folding free energy (ΔG_f) was calculated, as shown in Table 3, according to the following simple scheme:



since the sedimentation equilibrium and sedimentation velocity measurements of FP1 indicated that FP1 molecular species at 3mM FP1 concentration were mainly the oligomer close to the monomer and the oligomer consisting of about 15-20 monomers.

Secondly, the dissolution free energy of solute p (ΔG_{dissol}), which was defined as

$$\Delta G_{\text{dissol}} = -RT \ln S_p \quad [7],$$

where R was gas constant, T was absolute temperature, and S_p was the solubility of p . Since ΔG_{dissol} was related to the solubility of p , ΔG_{dissol} was depended on polarity, hydrophobicity, and the net charge of a protein solute^{19,28,29}, which positively increased with a decrease in pH. Strictly, equation [5] and [7] indicate that ΔG_{dissol} is dependent on a difference between a chemical potential of the solid phase, $\mu_{p(s)}$, and a chemical potential of solute p in an ideal aqueous solution at a standard solute concentration, $\mu_{p(\text{sol})}^0$, thus, the dissolution free energy could be regarded as a free energy of transfer the solute from the solid phase to the aqueous solution. ΔG_{dissol} of FPs at each pH related to the structural stability measurement were calculated by using r_p , $(\mu_{p(s)} - \mu_{p(\text{sol})}^0) / RT$ obtained by the fitting, and the net charge of FPs (Fig.4 B and Table 3). By comparison with ΔG_{dissol} , since the hydration potential (ΔG_{hyd}) is the free energy of transfer of a protein solute molecule from the gaseous phase into water, ΔG_{hyd} also depended on polarity of the solute and charge number of the ionized form, in which

protonated cations increased and deprotonated anions decreased with a decrease in pH ²¹. Strictly, ΔG_{hyd} is defined as a difference between $\mu_{p(\text{sol})}^0$ and a standard chemical potential of the gaseous phase of solute p , $\mu_{p(\text{g})}^0$:

$$\Delta G_{\text{hyd}} = \mu_{p(\text{sol})}^0 - \mu_{p(\text{g})}^0 \quad [8].$$

ΔG_{hyd} of FPs at each pH related to the structural stability measurement were calculated by using hydration potentials of amino acid side chains and the backbone ^{21,22} in addition to pK values for the amino acid side chains, $\alpha\text{-COOH}$ and $\alpha\text{-NH}^{3+}$ termini ³⁰ (Table 3).

Plotting of the structural stability against ΔG_{hyd} or ΔG_{dissol} showed that the structural stability was more strongly correlated with the dissolution free energy ($r = -0.86$; Fig.7 B) than with the hydration potential ($r = -0.70$; Fig.7 A). Then, transverse dispersions of the plots represented by open circles, which were at close pH, in Fig.7 A and B mainly reflected changes of ΔG_{hyd} and ΔG_{dissol} by the amino acid replacements of FPs, respectively, since ΔG_{hyd} and ΔG_{dissol} at same pH were dependent on the molecular properties without the electric charge. The open circle plots of FPs showed that these structural stabilities were almost identical while change of ΔG_{dissol} by the replacements of Phe26Tyr or Val28Ser was lower than that of ΔG_{hyd} , resulting in a difference of correlation coefficients shown in Fig.7 A and B. Furthermore, there was a strong relationship between the structural stability and the hydration free energy, indicating that destabilization of the dissolved state of a protein resulted in stabilization of the protein structure. Then, the more strong correlation with the dissolution free energy than with the hydration free energy indicated that destabilization of the dissolved state compared to the solid state resulted in stabilization of the protein structure, that is, the more insoluble a protein, the more stable the structural stability. Therefore, these

experimental results suggest that the structural stability is determined by protein solubility rather than degree of hydration.

Protein segment property necessary to induce formation of a soft-packed structure.

When the hydrophobic peptide fragment of FIVVALG was added to the C-terminus of IP consisting of 25 residues, which was unfolded, formation of the whole protein structure, FP1, was induced. The FP1 structure with the local structure including α -helix was stabilized by the intermolecular hydrophobic interactions, indicating that the added peptide fragment conferred the ability to form a low-specific structure, or soft-packed structure. Additionally, Phe26Tyr replacement of the added peptide fragment yielded the identical structural stability of the whole protein, FP2, compared to that of FP1, resulted from the fact that the protein solubility was not largely changed by the replacement. On the other hand, Val28Ser in addition to Phe26Tyr replacement yielded the drastic decrease in the structural stability of the whole protein, FP3, compared to that of FP1 or FP2, resulted from the fact that the protein solubility of FP3 largely increased by their replacements. A decrease of 1.5 kcal/mol in dissolution free energy by the replacements of Phe26Tyr and Val28Ser resulted in a decrease of ~ 5 kcal/mol of the structural stability (Fig.7 B), indicating that hydrophilic replacements of two out of seven hydrophobic residues in the peptide fragment deprived the whole protein of the ability to form the soft-packed structure. That is, all of or six residues in the peptide fragment should be hydrophobic to induce formation of soft-packed structure. The fact that the protein segment consisting of more than six correctly chosen amino acids conferred the ability to form the soft-packed structure to the other segment consisting of 25 randomly chosen amino acids verified that all of amino acids composing a protein needed not to

be correctly chosen to form a low-specific structure throughout the length of the molecule.

The increase in hydrophobicity resulted in the drastic increase in the structural stability, suggesting that the hydrophobic amino acids of the peptide fragment are buried in the protein interior⁹⁻¹¹. Furthermore, contrary to the notion that hydrophobicity is the dominant force of protein folding¹⁴, this validation suggests that only a significant increase in hydrophobicity of a protein segment in the primary sequence results in creation of a low-specific structure throughout the length of the molecule. Recently, Hecht and co-workers have created a four-helix-bundle protein consisting of 102 residues from a binary hydrophobic-polar patterned library, suggesting that sequences that fold into well-packed structures are not extremely rare³¹. For guess of probabilities of forming the well-packed structure and the soft-packed structure observed in FP1, a probability that a protein consisting of 32 amino acid residues has the four-helix-bundle structure, which is used as an example of well-packed structures, is calculated to be $1/2^{32} \approx 1/10^{10}$ while a probability that the protein has a soft-packed structure is calculated to be $1/2^6 \approx 1/10^2$ by assuming that a probability that an amino acid is polar or hydrophobic is 1/2. It is thus suggested that soft-packed structures including oligomers stabilized by intermolecular interactions can be formed much more frequently than well-packed structures. This especially high frequency is attributed to the fact that a soluble unfolded protein easily becomes insoluble by adding a hydrophobic peptide fragment to the C-terminus of the original protein. Actually, the solubility of IP is higher than 2.2mM at pH 7.1 (data not shown) though that of FP1 becomes $\sim 10\mu\text{M}$ (Fig4 A) by adding seven hydrophobic residues, indicating that addition of hydrophobic amino acids equivalent to $6/25 \approx 1/4$ length of a soluble unfolded protein consisting of

unbiased amino acids are needed to make the whole protein insoluble. This simple mechanism to stabilize soft-packed structures might be the reason that natural proteins could have evolved to well-packed structures in the possible primary sequence space.

Acknowledgements. We thank Miyo Sakai (Institute for Protein Research, Osaka University) for ultracentrifuge measurements.

References

1. Carlsson, U. & Jonsson, B. H. Folding of beta-sheet proteins. *Curr Opin Struct Biol* 5, 482-7 (1995).
2. Chakrabartty, A. & Baldwin, R. L. Stability of alpha-helices. *Adv Protein Chem* 46, 141-76 (1995).
3. Jackson, S. E. How do small single-domain proteins fold? *Fold Des* 3, R81-91 (1998).
4. Kuhlman, B. & Baker, D. Exploring folding free energy landscapes using computational protein design. *Curr Opin Struct Biol* 14, 89-95 (2004).
5. Ponder, J. W. & Richards, F. M. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J Mol Biol* 193, 775-91 (1987).
6. Dahiyat, B. I. & Mayo, S. L. De novo protein design: fully automated sequence selection. *Science* 278, 82-7 (1997).
7. Desjarlais, J. R. & Handel, T. M. De novo design of the hydrophobic cores of proteins. *Protein Sci* 4, 2006-18 (1995).
8. Isogai, Y., Ito, Y., Ikeya, T., Shiro, Y. & Ota, M. Design of lambda Cro fold: solution structure of a monomeric variant of the de novo protein. *J Mol Biol* 354, 801-14 (2005).
9. Filikov, A. V. et al. Computational stabilization of human growth hormone. *Protein Sci* 11, 1452-61 (2002).
10. Dantas, G., Kuhlman, B., Callender, D., Wong, M. & Baker, D. A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. *J Mol Biol* 332, 449-60 (2003).
11. Malakauskas, S. M. & Mayo, S. L. Design, structure and stability of a hyperthermophilic protein variant. *Nat Struct Biol* 5, 470-5 (1998).

12. Kuhlman, B. et al. Design of a novel globular protein fold with atomic-level accuracy. *Science* 302, 1364-8 (2003).
13. Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T. & Kim, P. S. High-resolution protein design with backbone freedom. *Science* 282, 1462-7 (1998).
14. Dill, K. A. Dominant forces in protein folding. *Biochemistry* 29, 7133-55 (1990).
15. Riener, C. K., Kada, G. & Gruber, H. J. Quick measurement of protein sulfhydryls with Ellman's reagent and with 4,4'-dithiodipyridine. *Anal Bioanal Chem* 373, 266-76 (2002).
16. Piotto, M., Saudek, V. & Sklenar, V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* 2, 661-5 (1992).
17. Wuthrich, K. (Wiley, New York, 1986).
18. Guntert, P., Mumenthaler, C. & Wuthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* 273, 283-98 (1997).
19. Tanford, C. *Physical chemistry of macromolecules*. (John Wiley and Sons, Inc., New York., 1961).
20. Narayan, V. A., Kriwacki, R. W. & Caradonna, J. P. Structures of zinc finger domains from transcription factor Sp1. Insights into sequence-specific protein-DNA recognition. *J Biol Chem* 272, 7801-9 (1997).
21. Wolfenden, R. V., Cullis, P. M. & Southgate, C. C. Water, protein folding, and the genetic code. *Science* 206, 575-7 (1979).
22. Privalov, P. L. & Makhatadze, G. I. Contribution of hydration to protein folding thermodynamics. II. The entropy and Gibbs energy of hydration. *J Mol Biol* 232, 660-79 (1993).
23. Wouters, M. A. & Curmi, P. M. An analysis of side chain interactions and pair correlations within antiparallel beta-sheets: the differences between backbone hydrogen-bonded and non-hydrogen-bonded residue pairs. *Proteins* 22, 119-31 (1995).
24. Gohon, Y. et al. Partial specific volume and solvent interactions of amphilipol A8-35. *Anal Biochem* 334, 318-34 (2004).
25. Stejskal, E. O. & Tanner, J. E. Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *J. Chem. Phys.* 42, 288-92 (1965).
26. Dingley, A. J. et al. Measuring protein self-association using pulsed-field-gradient NMR spectroscopy: application to myosin light chain 2. *J Biomol NMR* 6, 321-8 (1995).
27. Araki, M. & Tamura, A. Transformation of an alpha-helix peptide into a beta-hairpin induced by addition of a fragment results in creation of a coexisting state. *Proteins*

- (2006).
28. Ries-Kautt, M. & Ducruix, A. Inferences drawn from physicochemical studies of crystallogenes and precrystalline state. *Methods Enzymol.* 276, 23-59 (1997).
 29. Shaw, K. L., Grimsley, G. R., Yakovlev, G. I., Makarov, A. A. & Pace, C. N. The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. *Protein Sci* 10, 1206-15 (2001).
 30. Edsall, J. T. & Wyman, J. *Biophysical Chemistry, ch.8* (Academic Press, 1958).
 31. Wei, Y., Kim, S., Fela, D., Baum, J. & Hecht, M. H. Solution structure of a de novo protein from a designed combinatorial library. *Proc Natl Acad Sci USA* 100, 13270-3 (2003).
 32. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8, 477-486 (1996).

Table 1. Sequences of Sp1f3, IP, and FPs.

Sp1f3:	KKFACPECPK¹⁰	RFMRSDHLSK²⁰	HIKTHQNKK	
IP:	YAFACPACPK	RFMRSDALSK	HIKTA	
FP1:	YAFACPACPK	RFMRSDALSK	HIKTAFIVVA³⁰	LG
FP2:	YAFACPACPK	RFMRSDALSK	HIKTAYIVVA	LG
FP3:	YAFACPACPK	RFMRSDALSK	HIKTAYISVA	LG

Gly was inserted in the C terminus of FPs due to the resin of the chemical synthesis.

Table 2. Structure statistics of 10 FP1 structures

<i>Number of distance restraints</i>	
Intraresidue	150
Short-range ($ i-j = 1$ residues)	118
Medium-range ($ i-j = 2$ to 4 residues)	74
Long-range ($ i-j > 4$ residues)	(52) *
<i>Number of torsion angle restraints</i>	
ϕ	25
<i>Geometric statistics</i>	
R.m.s deviation from the mean structure (Å)	
Backbone atoms (residues 3-30)	2.79 ± 0.71
All heavy atoms (residues3-30)	3.51 ± 0.81
<i>Ramachandran analysis **</i>	
Most favored regions (%)	42.1
Additional allowed regions (%)	37.5
Generously allowed regions (%)	18.9
Disallowed regions (%)	1.4

* Long-range distance restraints were not used in the structure calculation (See the text).

** Ramachandran analysis was evaluated by using the program PROCHECK ³².

Table 3. The pH dependence property of ΔG_f , ΔG_{dissol} , and ΔG_{hyd} of FPs

	pH	ΔG_f (kcal/mol)	ΔG_{dissol} (kcal/mol)	ΔG_{hyd} (kcal/mol)
FP1	2.67	-44.4 ± 0.2	-3.86	-500.4
	2.85	-46.5 ± 0.2	-3.24	-498.9
	3.05	-49.6 ± 0.3	-2.59	-497.2
	3.27	-55.1 ± 0.4	-1.90	-495.4
	3.52	-54.1 ± 0.3	-1.10	-493.4
	3.91	-67.4 ± 2.7	0.169	-490.4
	4.03	-59.2 ± 0.8	0.545	-489.5
FP2	3.00	-47.9 ± 0.3	-2.93	-502.9
	3.17	-48.1 ± 0.3	-2.40	-501.5
	3.31	-50.6 ± 0.4	-1.96	-500.4
	3.86	-49.5 ± 0.3	-0.180	-496.1
	4.16	-58.6 ± 1.0	0.740	-493.9
FP3	3.87	-49.0 ± 0.4	-1.34	-503.1
	4.20	-62.0 ± 0.6	-0.342	-500.6
	4.47	-57.3 ± 0.4	0.315	-498.7
	4.82	-59.6 ± 1.3	0.958	-496.2
	5.65	-67.0	2.43	-490.7

Errors of pH and ΔG_{dissol} were less than 0.02 and 0.1, respectively.

A error of ΔG_f of FP3 at pH5.65 could not be obtained since only the spectrum at mixing time of 300ms was analyzed because of an increase in the aggregation rate.

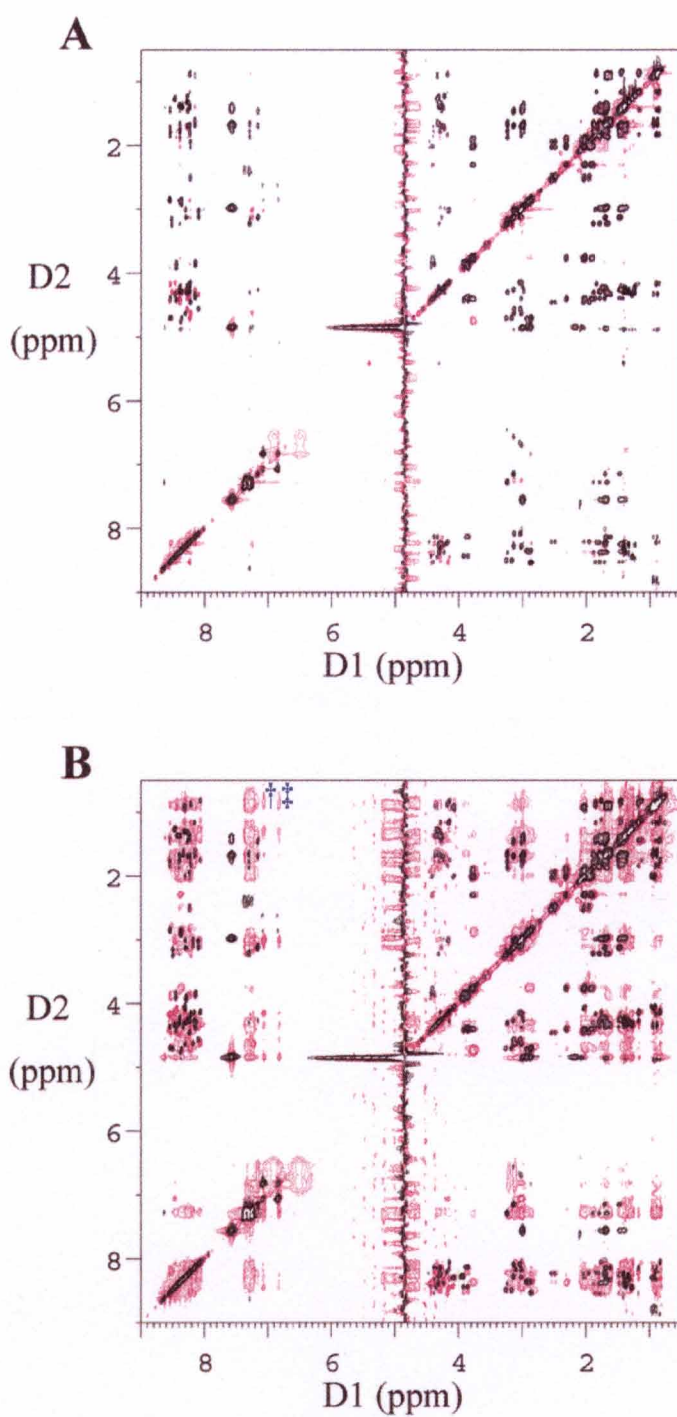


Fig.1 Two-dimensional ^1H - ^1H NMR spectra of IP (A) and FP (B). NOESY and TOCSY spectra were drawn in red and black, respectively. $\text{Y1C}^{\delta}\text{H}$ - $\text{I27C}^{\gamma}\text{H}$ and $\text{Y1C}^{\epsilon}\text{H}$ - $\text{I27C}^{\gamma}\text{H}$ cross-peaks are marked with *dagger* and *double-dagger*, respectively.

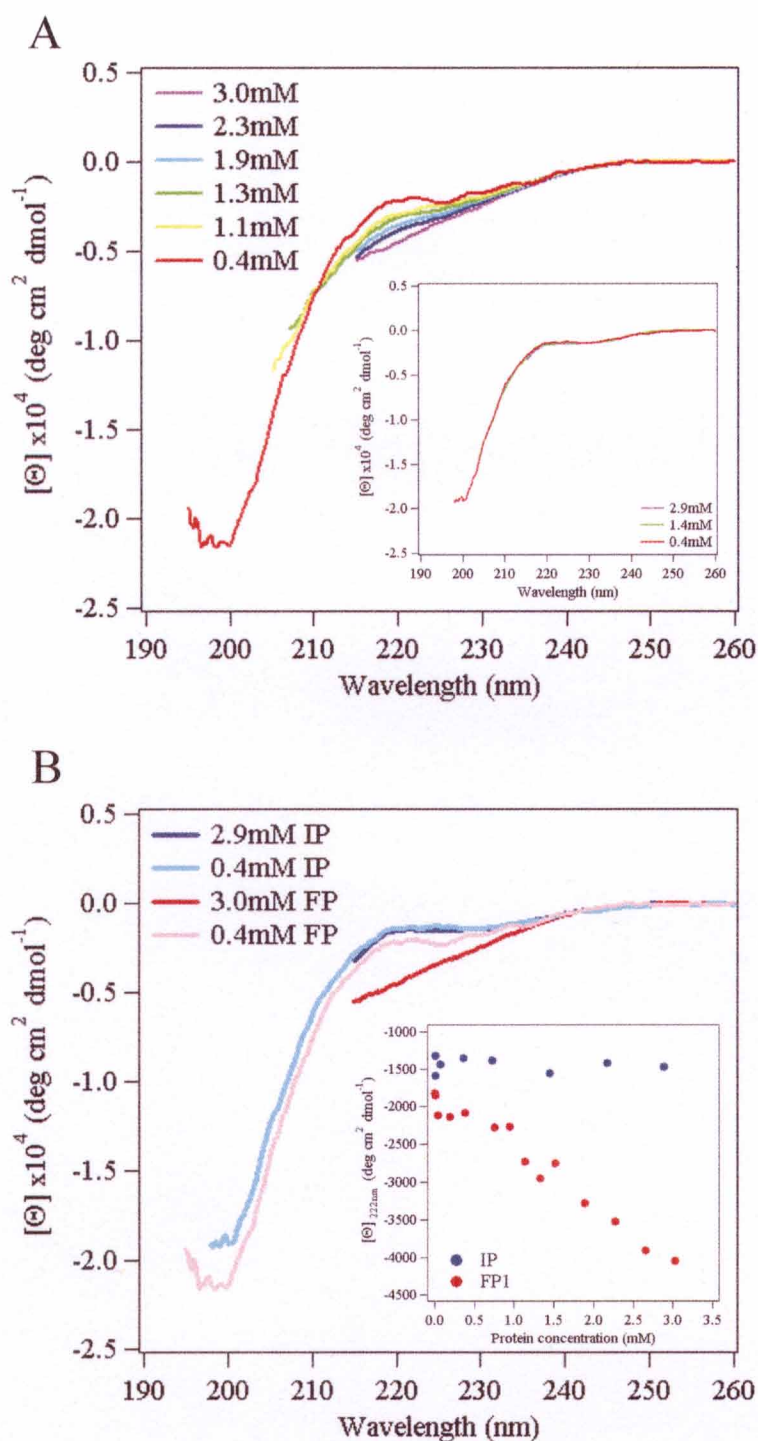


Fig.2 The protein concentration dependence of far-UV CD spectra of IP and FP1. (A) Spectra of FP1. (Inset) Spectra of IP. (B) Spectra of IP and FP1. (Inset) $[\Theta]$ at 222nm of IP and FP1. $[\Theta]$ was represented as molar ellipticity per residue.

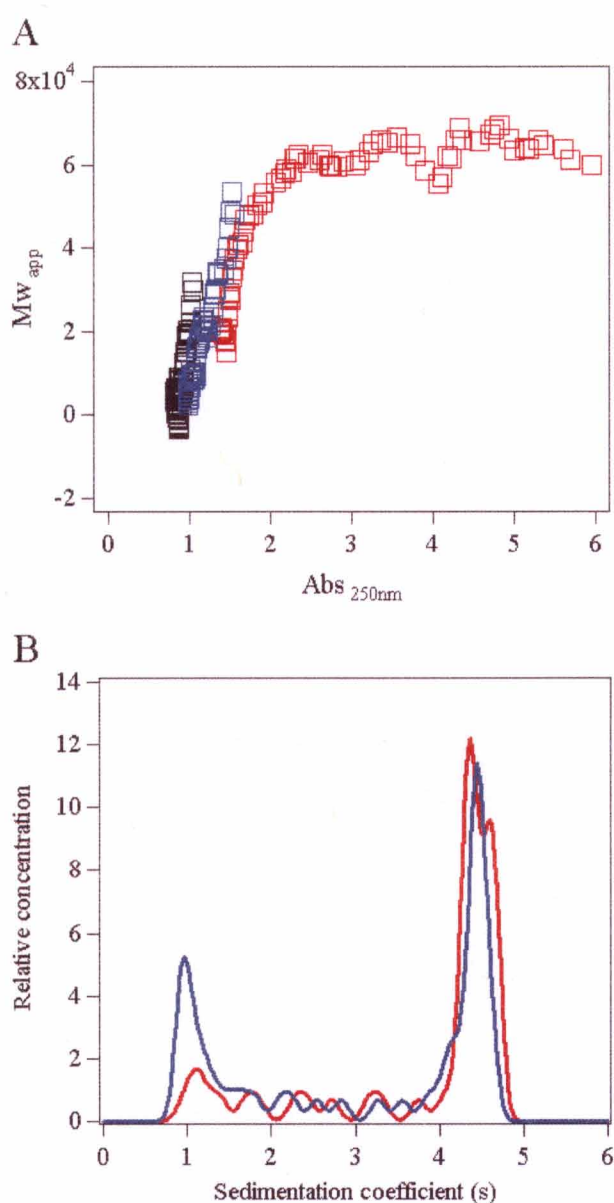


Fig.3 Sedimentation equilibrium and sedimentation velocity measurements of FP1. (A) The distribution of apparent molecular weight (Mw_{app}) against the location in the cell obtained from the sedimentation equilibrium measurements at 0.9mM (black), 1.5mM (blue), and 3.0mM (red) protein concentrations. The apparent molecular weight was calculated at respective points in the cell, *i.e.* the higher the absorbance at 250nm (Abs_{250nm}), the closer to the bottom of the cell. (B) The distribution of sedimentation coefficients obtained from the sedimentation velocity measurements at 2.3mM (blue), and 3.0mM (red) protein concentrations.

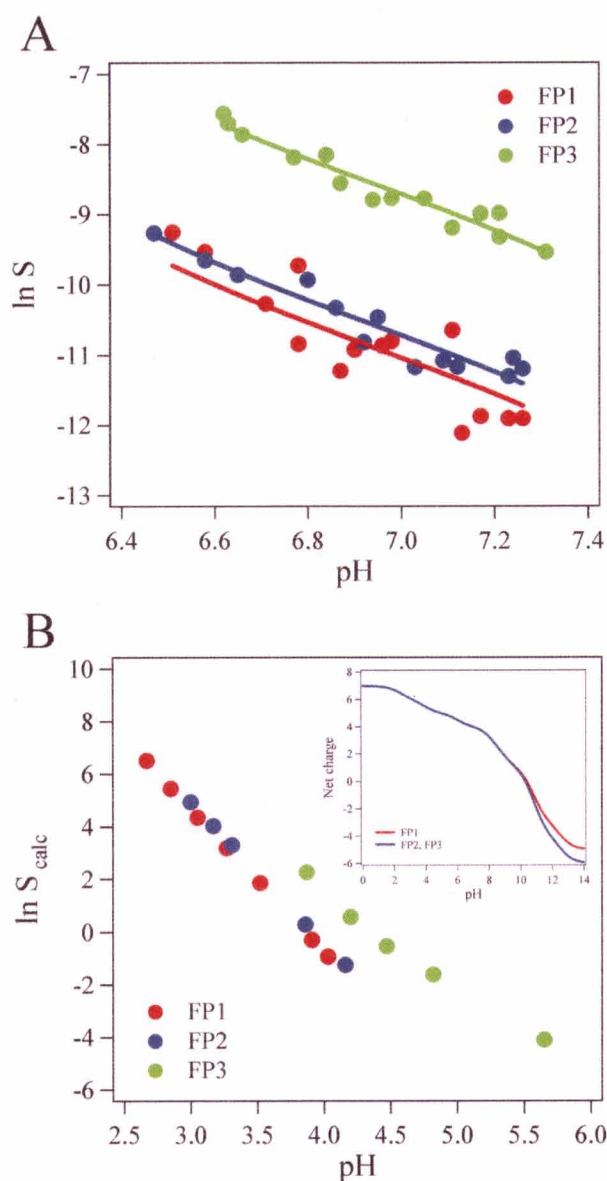


Fig.4 The pH dependence of the solubilities of FPs (A) Experimentally obtained plots of the natural logarithm of the solubility, S (mol/l), in the pH range of 6.4-7.4. (B) The solubilities calculated by using $r_p=384\text{\AA}$, $(\mu_{p(s)} - \mu_{p(sol)}^0) / RT$ of FP1, FP2, and FP3 were -22.8 ± 0.1 , -22.5 ± 0.1 , and -20.5 ± 0.0 , respectively, in the pH range of 2.8-6.0. Errors of $\ln S_{\text{calc}}$ of FPs were smaller than 0.1. (Inset) Net charge curves of FPs calculated by using pK values for the amino acid side chains, the $\alpha\text{-COOH}$ and the $\alpha\text{-NH}^{3+}$ termini: Tyr = 10.9, Cys = 8.3, Lys = 10.8, Arg = 12.5, Asp = 3.9, His = 6.0, NH^{3+} of the N terminus = 9.1, and COOH of the C terminus = 2.4³⁰.

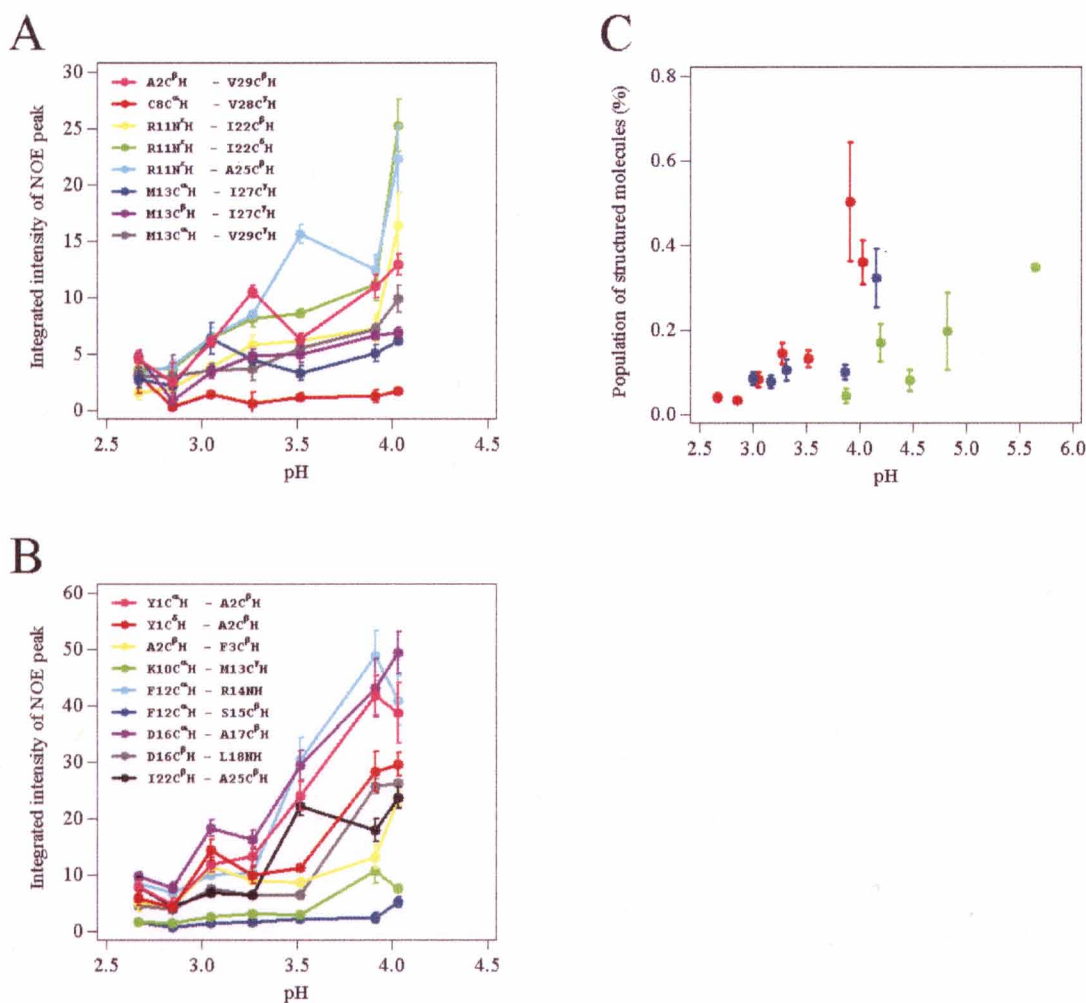


Fig.5 The pH dependence of the structural stabilities of FPs. (A) Integrated intensities of long-range NOE peaks of FP1. (B) Integrated intensities of short and medium-range NOE peaks of FP1. (C) Averaged population of structured molecules (%) of FP1 (red), FP2 (blue), and FP3 (green).

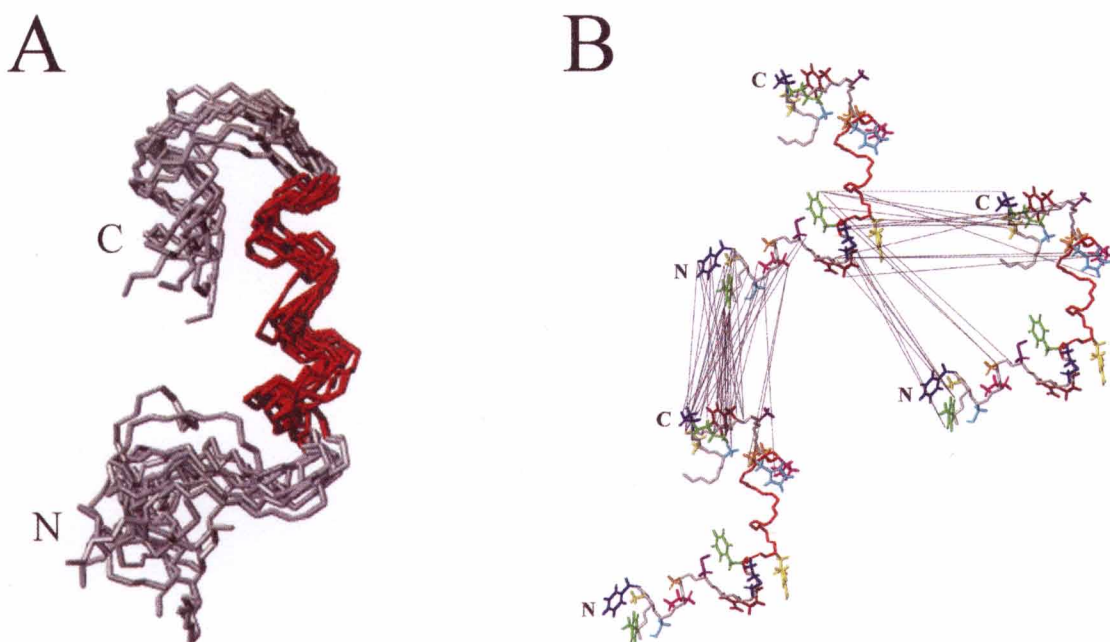


Fig.6 NMR structures of FP1. (A) Backbone traces of the 10 best structures. Backbones of residues 12-22, which adopted α helix, were drawn in red. (B) Schematic diagram of long-range interactions between FP1 molecules. Long-range interactions were represented by black lines between the lowest energy structures of FP1, of which the side chain of Tyr1 Phe12 Ile27 (green), Phe3 Met13 Val28 (blue), Ala4 Arg14 Val29 (yellow), Cys5 His21 Ala30 (cyan), Pro6 Ile22 (magenta), Ala7 Lys23 (orange), Cys8 Ala25 (purple), and Arg11 Phe26 (brown) were shown. Additionally, NOE peaks including aromatic protons of Phe3 and Phe12 of FP1 were not clearly separated since chemical shifts of aromatic protons of Phe3 and Phe12 were close to those of Phe26. Therefore, long-range interactions including aromatic protons of Phe3 and Phe12 of FP2, which was the variant Phe26Tyr of FP1, were added.

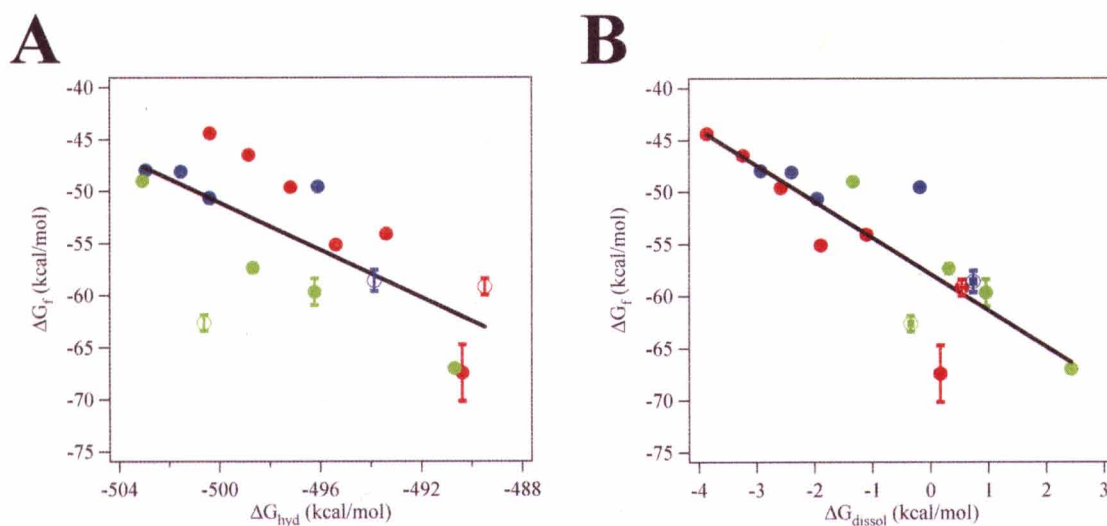


Fig.7 The relationships between structural stability and molecular properties. The correlation between the folding free energy (ΔG_f) and (A) the hydration potential (ΔG_{hyd}), which was defined as the free energy of transfer of a protein solute molecule from the gaseous phase into water, or (B) the dissolution free energy ($\Delta G_{dissol} = -RT \ln S$) for FP1 (red), FP2 (blue), and FP3 (green). Plots of FP1 at pH 4.0, FP2 at pH4.2, and FP3 at pH4.2 were represented by open circles. The lines represented linear fits with correlation coefficients of -0.70 and -0.86, respectively. For the calculation of ΔG_{hyd} of FPs, hydration potentials of 18 amino acid side chains excluding Pro and Arg were taken from the values measured by Wolfenden et al.²¹. Those of Pro, Arg side chains and the backbone were taken from the values measured by Privalov et al.²². pK values of the amino acid side chains, the α -COOH and the α -NH³⁺ termini were taken from the values represented in the legend of Fig.4³⁰. The error bars of ΔG_{dissol} were less than 0.1.

Acknowledgements

I wish to express my gratitude to Dr. Atsuo Tamura in Graduate School of Science and Technology, Kobe University on the point that he gave me education that develops individuality of academic. Additionally, I would like to express my gratitude to Prof. Motonari Tsubaki, Prof. Fumio Hayashi, and Prof. Keisuke Tominaga in Faculty of Science, Kobe University, on the point that they gave me advices for writing my doctoral dissertation. Finally, I have my family to thank for helping me.

Publication lists

Mitsugu Araki and Atsuo Tamura: Transformation of an α -helix peptide into a β -hairpin induced by addition of a fragment results in creation of a coexisting state. *Proteins: Structure, Function, and Bioinformatics*, 66, 860-868, 2007

Mitsugu Araki and Atsuo Tamura: Protein segment that drastically decreases the solubility induces formation of the whole protein structure. (in preparation)