Key Elements for Protein Foldability Revealed by a Combinatorial Approach among Similarly Folded but Distantly Related Proteins†

Sayuri Morimoto‡ and Atsuo Tamura*,‡,§

PRESTO, Japan Science and Technology Corporation, and Graduate School of Science and Technology, Kobe University, Nada, Kobe 657-8501, Japan

Received November 18, 2003; Revised Manuscript Received February 26, 2004

ABSTRACT: We have determined the key regions for protein foldability by creating multiple crossover libraries from two proteins that share similar fold but have low sequence identity and differ significantly in stability. One protein is the propeptide of a serine protease, subtilisin BPN′, and the other is *Pleurotus ostreatus* proteinase A inhibitor 1 (POIA1). The propeptide has a compact structure when complexed with subtilisin but is unstructured when isolated, whereas POIA1 takes a stable structure. We selected four of the conserved amino acid residues for the boundaries of crossover sites and utilized these residues to make same cohesive-ends to assemble synthetic DNA fragments. Each segment has one or two secondary structure units, and the interchange of these structural elements produces 32 ($= 2^5$) combinations, including the propeptide and POIA1. The stability of these mutants was first screened by formation of turbid zones on skim milk plates containing subtilisin BPN′. It was shown that six variants were foldable and structural units necessary for folding were identified. Further fragmentation and recombination of these mutants (the "multisection" method) revealed that two interactions between secondary structures are important; one is interaction between the loop-α1 and β2-turn-β3, and the other is hydrophobic interaction between the adjoining β1 and β4 strands. We were also able to specify the significant amino acid combinations for tolerance to proteolysis. These combinatorial methods not only elucidate how domains can be interchanged to make the whole protein foldable but also extract essential regions for the function, which is correlated with the instability of the molecule.

Combinatorial experiments can readily generate sequence ensembles, and diversity of these additional sequences will be at the control of the researcher (*1*−*4*) and can promote acquisition of novel properties in proteins. In addition to DNA shuffling and its variants (*5*−*7*), which recombine only closely related sequences, recombination protocols capable of freely exchanging genetic diversity without sequence identity limitations have been extensively developed in laboratory evolution studies (*8*−*10*). One of these approaches is assembly of structural elements called "building blocks", which has been employed for the structure−function study (*11*−*14*). We have applied this type of building block approach in a different manner and created crossovers between structurally homologous proteins to study not only the function but also the protein foldability. Two proteins were chosen for making the hybrids; one is a natively unfolded protein, the propeptide of subtilisin BPN′, and the other is a folded protein, *Pleurotus ostreatus* proteinase A inhibitor 1 (POIA1).[1]

Subtilisin BPN′ is synthesized as preprosubtilisin, which consists of a 30 amino acid signal peptide, a 77 amino acid propeptide, and a 275 amino acid mature form of the enzyme (*15*, *16*). The propeptide of bacterial subtilisin is essential for facilitating proper folding of the mature form (*17*−*21*). Upon completion of folding, the propeptide is autoprocessed and remains bound to the active site of subtilisin as a temporary inhibitor (*22*−*24*). The subsequent autodegradation of the propeptide then leads to the release of the mature subtilisin (*25*, *26*). In the complex with the inactive subtilisin, the propeptide has a compact structure (*27*, *28*) (Figure 1A,B). When isolated, however, the propeptide becomes completely unstructured (*21*, *29*). Several attempts to have the isolated propeptide folded were made by accumulating point mutations on the basis of the X-ray crystal structure of the propeptide−subtilisin complex. One mutant with several amino acid substitutions including one disulfide cross-link was found to be foldable (*30*). Besides this, some mutants formed unexpected dimers at neutral pH (*31*−*33*) and other mutants were not fully folded (*34*). These results indicate that a significant number of mutations are required for formation of the tertiary structure in the isolated propeptide.

POIA1, which consists of 76 amino acid residues and is structured by itself (*35*, *36*), is related distantly to the propeptide of subtilisin BPN′ (18% amino acid sequence identity), although the secondary and the tertiary structure is almost identical to the propeptide in the complex with

† A.T. was financially supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture of Japan.

* To whom correspondence should be addressed. Tel: ++81-78-803-5692. Fax: ++81-78-803-5692. E-mail: atamura@kobe-u.ac.jp.

‡ Japan Science and Technology Corporation.

§ Kobe University.

[1] Abbreviations: POIA1, *Pleurotus ostreatus* proteinase A inhibitor 1; IPTG, isopropyl-β-D-thiogalactoside; GST, glutathione S-transferase; CD, circular dichroism; NMR, nuclear magnetic resonance.

10.1021/bi0302402 CCC: $27.50
Published on Web 05/05/2004

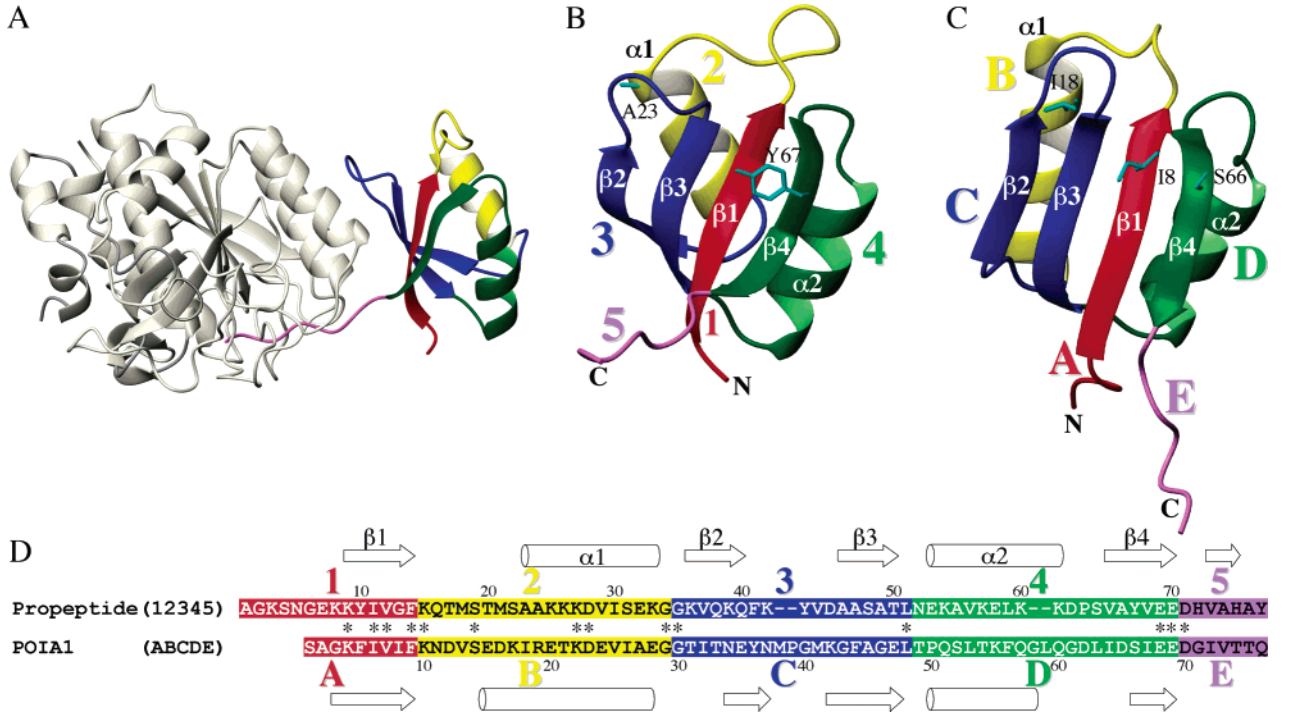

FIGURE 1: Structural comparison of the propeptide of subtilisin BPN′ and POIA1 depicted with five structural segments in different colors. Each segment is termed as 1−5 for the propeptide and A−E for POIA1. Panel A displays a ribbon diagram of the complex between mature subtilisin BPN′ (gray) and the propeptide (color) from the crystal structure (PDB code 1spb). Panel B shows the structure of the propeptide when coupled with subtilisin BPN′. Note that the propeptide is natively unstructured by itself. Panel C shows the structure of POIA1 from the solution structure (PDB code 1itp). The mutated side chains are displayed in cyan. All structures are displayed using MOLMOL (*43*). Panel D presents sequence alignment of the propeptide and POIA1 with the secondary structure units. Asterisks show identical amino acids between the propeptide and POIA1.

subtilisin BPN′ (*37*) (Figure 1C). Both the propeptide, when coupled with subtilisin, and POIA1 have the same (β-α-β) × 2 topology and are classified as ferredoxin-like fold by Structural Classification of Proteins (SCOP) (*38*), although they show different stability and functions. Thus, we take a strategy to make hybrid mutants of the two proteins to identify building blocks necessary for folding, since it is difficult to specify individual amino acid residues by merely comparing the structure of the two proteins. In this study, we have developed a simple way of generating libraries of chimeric proteins, which utilized conserved amino acids between the propeptide and POIA1 as crossover sites (Figure 1D). The advantage of this building block approach is that we can efficiently search a large sequence space that cannot be attained by just accumulating point mutations. The chimeragenesis can also be a quest for functional importance of the unstructured state, since the propeptide is a natively unstructured but functional protein.

MATERIALS AND METHODS

Materials. Synthetic DNA oligonucleotides were obtained from Invitrogen. DNA-modified enzyme, restriction enzyme, and isopropyl-β-D-thiogalactoside (IPTG) were purchased from Toyobo (Kyoto, Japan). A synthetic peptide substrate, Suc-Ala-Ala-Pro-Phe-MCA, was from Peptide Institute (Osaka, Japan). Type PVDF membrane filters (0.45 μm pore size) were obtained from Millipore. *Escherichia coli* strain BL21, expression vector pGEX-2X, affinity column for purifying glutathione S-transferase (GST) fusion proteins, and Superdex-75 gel-filtration column were purchased from Amersham Biosciences. The subtilisin BPN′ expression vector and the host strain *Bacillus subtilis* UOT0999 were gifts from Dr. Seiichi Taguchi (Meiji University, Tokyo, Japan).

Gene Construction. In designing the sequences of DNA fragments (Figure 2), we mainly employed the most frequent codons used in *E. coli*. After phosphorylation of all fragments except Pro1U, Pro5L, PO AU, and PO EL, upper and lower oligonucleotides of each segment were annealed. Then double-stranded DNA fragments were individually ligated by T4 DNA ligase to produce 32 combinations. The 32 variants of synthetic genes were inserted into the pGEX-2X expression vector, respectively, at the *Bam*HI and *Eco*RI restriction sites to yield pGEX-12345−pGEX-ABCDE, resulting in expression of fusion proteins connected with the C-terminus of GST through the thrombin recognition site. The oligonucleotides that were used to introduce the site-specific mutagenesis by overlap extension for additional mutations of the hybrid proteins have the following sequences (5′ to 3′): 2-F1, CTTCTTAGCAGCAGACATAGTCGACATAGTCTGTTTGAA; 2-F2, AGTTTCACGGATTTTGTCTTCCGACATAGTCTGTTTGAA; 2-F3, CTTCTTAGCAGCAGACATAGTAGAGACGTCGTTTTTGAA; 2-F4, AGTTTCACGGATTTTGTCTTCAGAGACGTCGTTTTTGAA; 2-F5, ACTATGTCTGCTGCTAAGAAGAAAGACGTGATATCTGAA; 2-F6, ACTATGTCTGCTGCTAAGAAGAAAGACGAAGTTATCGCT; 2-F7, GAAGACAAAATCCGTGAAACTAAAGACGTGATATCTGAA; 2-F8, GAAGACAAAATCCGTGAAACTAAAGACGAAGTTATCGCT; I8G-L, TTTGAAACCAACGATGAATTTACCAGC; I8G-B-U, CATCGTTGGTTTCAAAAACGACGTCTCT; D-S66Y-L, TTCAATGTAGTCGATCAGGTCACC; D-S66Y-U, CT-

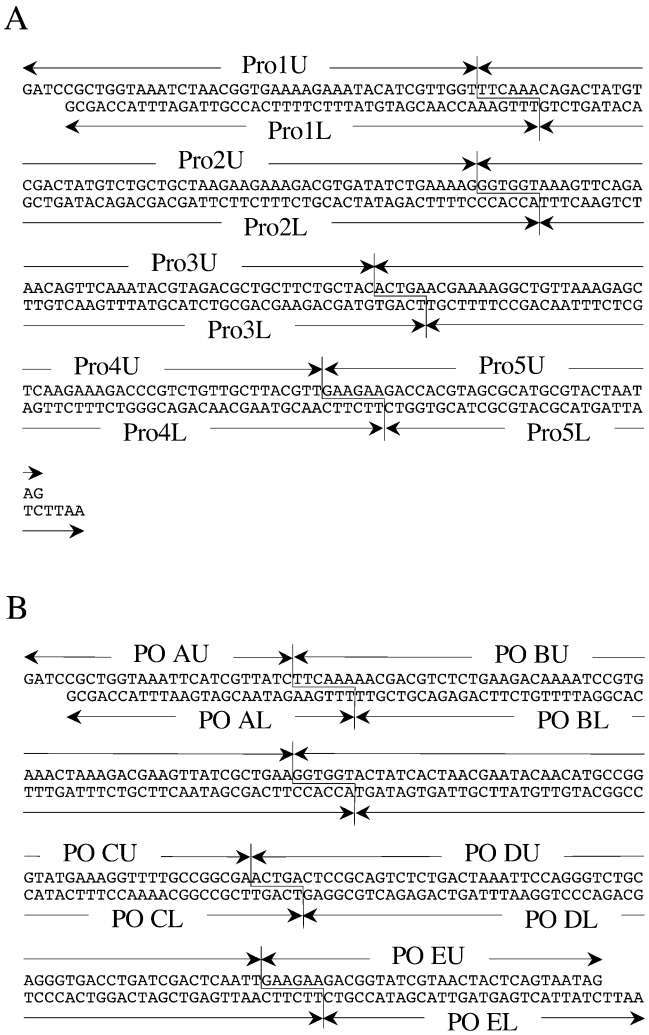

FIGURE 2: DNA sequences of the propeptide of subtilisin BPN′ (A) and POIA1 (B). Arrows above and below the sequence show the fragments of the synthetic oligonucleotides for hybrid construction. Note that one or two additional amino acids are located at the N-termini of the hybrids since the C-terminus of the thrombin recognition site remains.

GATCGACTACATTGAAGAAGAC; 1-G13I-L, TTTGAAGATAACGATGTATTTCTTTTC; G13I-B-U, CGTTATCTTCAAAAACGACGTCTCT; 2-A23I-U, ACTATGTCGACTATGTCTATCGCTAAGAAGAAAGACGTGA. The mutations were subsequently confirmed through DNA sequencing performed on a 310-Genetic Analyzer (Applied Biosystems Inc.).

In Vivo Selection. An *E. coli* strain, BL21, was individually transformed with pGEX-12345−pGEX-ABCDE. Each transformant expressing the GST fusion protein was then streaked separately on a membrane filter onto a Luria-Bertani (LB) agar plate containing 50 μg/mL of ampicillin and incubated at 37 °C for 18 h (Figure 3, left panel). Next, the filter was peeled off and placed on the surface of a skim milk−subtilisin BPN′ top agar plate. The skim milk−subtilisin BPN′ top agar plate was prepared by mixing 0.7% agarose, 5% skim milk, and subtilisin BPN′ (final concentration, 84 nM) in 9 mL of 0.1 M Tris-HCl (pH 8.0) at 50 °C and pouring onto a LB agar plate containing ampicillin. After incubation for 10 h at 25 °C, the filter was peeled off and the turbid areas of skim milk on the top agar plate (Figure 3, right panel) were monitored by visual inspection. Although

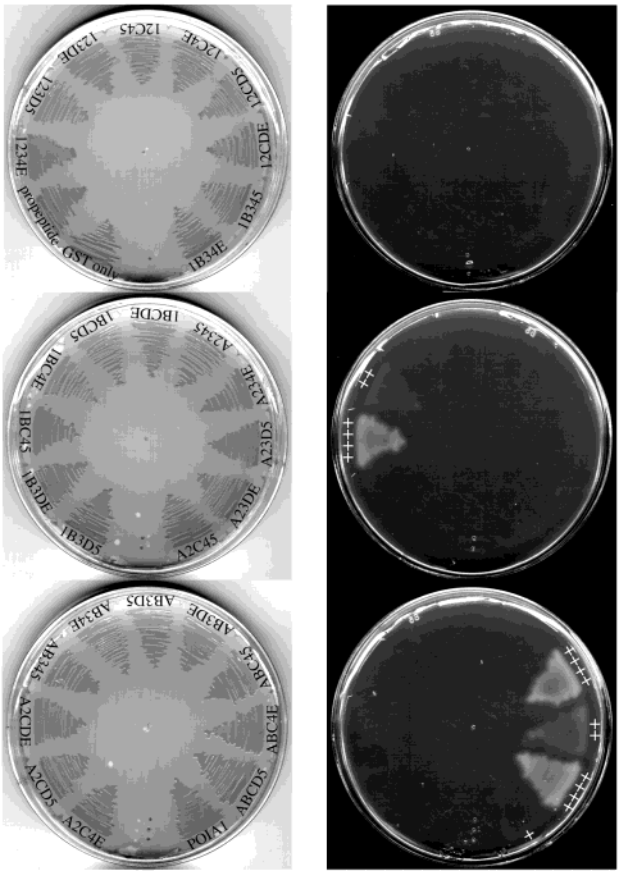

FIGURE 3: In vivo selection: (left) streaked *E. coli* strains individually transformed with pGEX-12345−pGEX-ABCDE on the membrane filter; (right) detection of turbidity. The filters in the left panel were peeled off and then located on skim milk−subtilisin BPN′ top agar plates for 10 h at 25 °C. After peeling off the filters, turbidity of each mutant was evaluated by visual inspection.

there is no secretion signal in the expression vector and thus the protein is expressed in the cytoplasm of *E. coli*, we confirmed that 0.17−0.22% of expressed GST fusion protein leaked out from the cells to the medium, presumably from destroyed host cells, during incubation for 10 h at 25 °C. These small amounts of proteins are sufficient to be detected as inhibitors in this assay system.

Expression and Purification of Proteins. Each colony containing the recombinant expression vector was grown in Terrific Broth supplemented with 50 μg/mL ampicillin. When the absorbance at 610 nm was around 0.6, protein expression was induced by adding IPTG (final concentration 1 mM). After 2 h at 25 °C, the cells were harvested by centrifugation, resuspended in PBS, and sonicated at 4 °C. Then the lysate was centrifuged to remove precipitate and applied to the affinity column at 4 °C. The eluted GST fusion protein was cleaved by thrombin for 16 h at 25 °C. Each hybrid mutant was then purified by gel-filtration chromatography equilibrated with PBS. It should be noted that additional amino acids are located at the N-termini of the hybrids in the propeptide (Gly-Ser) and POIA1 (Gly) because the C-terminus of the thrombin recognition site remains. The concentration of hybrid protein was determined by measuring the absorbance at 280 nm in 6 M guanidine hydrochloride (*39*). Subtilisin BPN′ was purified from a culture of *B. subtilis* UOT0999 described by Taguchi et al. (*40*).

Circular Dichroism (CD) and Nuclear Magnetic Resonance (NMR) Measurements. CD spectra were recorded on a Jasco J-720 spectropolarimeter (Tokyo, Japan). For far-UV CD spectra, a 1 mm path length cell was used with a 1 nm bandwidth. Eight scans were accumulated to obtain the spectra from 200 to 250 nm with an increment of 0.2 nm at 25 °C in PBS. To study the thermal stability of the proteins, the temperature of the sample in 25 mM sodium phosphate buffer at pH 7.0 was increased from 0 to 84 °C at 2 °C intervals with a temperature controller, PTC-348w (Jasco). The thermal melting curves were drawn on the basis of the ellipticity at 220 nm. ^{1}H NMR spectra were recorded with a Bruker DMX-750 spectrometer in 25 mM sodium phosphate (pH 7.0) at 25 °C in 99.9% D_2O.

Measurement of Inhibitory Activities and Tolerance to Proteolysis. Inhibition constants for the interactions of the hybrids with subtilisin BPN′ were determined from the equilibrium rates of cleavage of a fluorogenic substrate by the enzyme at different concentrations of the inhibitor. The hydrolysis of methyl coumarylamide substrate was monitored every 15 s using a SOFTmax PRO microplate fluorometer (Molecular Devices) at an excitation wavelength of 355 nm and an emission wavelength of 460 nm. Subtilisin BPN′ was incubated at room temperature in assay buffer (0.1 M Tris-HCl (pH 8.0), 2 mM $CaCl_2$, 0.005% Tween-20), using 100 μM Suc-Ala-Ala-Pro-Phe-MCA as a substrate. The hybrid protein with various concentrations was added, and the succeeding reaction was monitored. The extent of substrate hydrolysis never exceeded 5%, confirming that initial rates were indeed detected without exhaustion of the substrate. The concentrations of mutants were more than 10 times of the enzyme concentrations that caused 20−80% enzyme inhibition. For the case of the tight binding (<0.1 nM), fluorescence at 460 nm was monitored every 1 s at 36 μM substrate at 25 °C with a spectrofluorometer, FP-6500 (Jasco). These data were fitted to a theoretical equation $[P] = v_S t + (v_0 - v_S)(1 - e^{-kt})/k$, where [P] presents the product concentration, v_0 and v_S are the initial and steady-state velocities, respectively, k is the rate constant, and t is time in seconds. Apparent inhibition constants, K_i(app), were calculated as the slope of the plot of $[I]/(1 - v_S/v_0)$ against v_0/v_S, in which [I] is the inhibitor concentration (*41*). At least five different concentrations of inhibitor were used for each determination. For measurement of the tolerance to proteolysis, subtilisin BPN′ at 50 nM was incubated at room temperature with a 15-fold molar excess of the mutated protein in assay buffer. Aliquots at fixed time intervals were mixed with the substrate (100 μM), and velocity of the hydrolysis was monitored as a change in fluorescence at 460 nm. Inhibitory activities (%) were calculated by the equation $(1 - v_i/v_0) \times 100$, where v_i and v_0 are the initial rate of enzymatic reaction with and without the inhibitor, respectively.

RESULTS AND DISCUSSION

Construction of Hybrid Proteins and in Vivo Selection. Four conserved regions with one-to-three amino acids sequence identity were chosen for segment boundaries of crossover (Figure 1D) and utilized as same cohesive ends to assemble the synthetic oligonucleotide fragments (Figure 2) on the ligation step. These conserved regions are dispersed, and they reside mainly in the edges of the secondary structure. One possible cause for the existence of the homologous regions between distantly related proteins is that the conserved residues are necessary for the similar function as inhibitors during evolution since both of the target enzymes, subtilisin and *P. ostreatus* proteinase A, belong to the serine protease family. These five segments of the propeptide were named 1−5, and those of POIA1 were termed A−E, respectively (Figure 1B−D). The interchange of these five structural elements produced 32 (= 2^5) combinations (Table 1), including the propeptide (12345) and POIA1 (ABCDE). Each mutated protein was expressed as a fusion protein connected with the C-terminal region of GST.

We have developed a new method for detecting difference in stability by observing inhibitory activities in vivo (Figure 3). This selection is based on the fact that the propeptide, which is unstable, acts as a temporary inhibitor and is degraded gradually by subtilisin added *in trans*, while stable mutants such as POIA1 (ABCDE) or ABCD5, the latter having especially high inhibitory activity, serve as long-lived inhibitors (*42*). After incubation of the propeptide-expressing transformant (Figure 3, upper left) on a skim milk−subtilisin BPN′ top agar plate for 10 hours at 25 °C, the white color of the skim milk changed to transparent (Figure 3, upper right), since subtilisin, which has broad specificity for proteolysis, degrades the propeptide and subsequently the casein in skim milk. On the other hand, the area of ABCD5-expressing transformant streaked on a filter (Figure 3, lower left) remained turbid (Figure 3, lower right), indicating tolerance to proteolytic activity of subtilisin. When only GST was expressed, the transformant did not inhibit the proteolytic activity (Figure 3, upper). Therefore, it is proven to be possible to identify whether a mutant is degradable by means of this plate assay system. Turbidity of the hybrids estimated as five grades by visual inspection has been summarized in Table 1. Five hybrid proteins (1BC45, 1BC4E, ABC45, ABC4E, and ABCD5) and POIA1 have tolerance to proteolysis and thus are classified as stable proteins. Their second and third segments are all derived from POIA1, or BC is included, indicating that the most important interactions for stability exist between the second and third elements. Among XBCXX chimeras, where X stands for either segment in derivation, there is a clear rule in maintaining stability. When these hybrids have the fourth segment of the propeptide (XBC4X), they are foldable no matter which the first segment is derived from. On the other hand, the hybrids with the fourth segment of POIA1 (XBCDX hybrids) can be folded with only POIA1-derived first segment (ABCDX). These results suggest that the second important interaction for maintaining a stable state exists between the first and fourth segments.

Structural Characterization of Expressed Hybrids and Its Relation to the in Vivo Screening. We attempted to purify all these hybrids independently and to employ them for structural studies. Although all the proteins migrated to bands at the expected molecular weights with a certain expression level in SDS−polyacrylamide gel electrophoresis (data not shown), eight of the hybrids having a POIA1-derived first segment and propeptide-derived second segment (A2XXX) were degraded by proteolysis in the course of affinity chromatography and could not be purified. The other 24 hybrids were assessed for their conformational stability after

Table 1: Segment Configuration of the Hybrid Proteins and Results of In Vivo and In Vitro Experiments[a]

Hybrid[b]	Segment distribution[c]	Turbidity[d]	Foldability[e]	Hybrid	Segment distribution	Turbidity	Foldability
Propeptide(12345)		–	U	A2345		–	ND
1234E		–	U	A234E		–	ND
123D5		–	U	A23D5		–	ND
123DE		–	U	A23DE		–	ND
12C45		–	U	A2C45		–	ND
12C4E		–	U	A2C4E		–	ND
12CD5		–	U	A2CD5		–	ND
12CDE		–	U	A2CDE		–	ND
1B345		–	U	AB345		–	U
1B34E		–	U	AB34E		–	U
1B3D5		–	U	AB3D5		–	U
1B3DE		–	U	AB3DE		–	U
1BC45		++++	F	ABC45		++++	F
1BC4E		++	F	ABC4E		++	F
1BCD5		–	PF	ABCD5		++++	F
1BCDE		–	PF	POIA1(ABCDE)		+	F
$1(2_1 2_2 B_3)$C45		–	U				
$1(2_1 B_2 2_3)$C45		+++	U				
$1(2_1 B_2 B_3)$C45		+++	U				
$1(B_1 2_2 2_3)$C45		–	U				
$1(B_1 2_2 B_3)$C45		+	U				
$1(B_1 B_2 2_3)$C45		++++	PF				
12C45-A23I		++	U				

[a] Constructed hybrid proteins and results on stability. [b] Subscripts represent the partitioned segments in the second element. [c] Shading blocks indicate segments derived from POIA1. An asterisk shows the point mutation site, Ala23Ile. [d] Turbidity of the hybrids estimated as five grades by visual inspection: from ++++ (identical to the whiteness of the skim milk) to − (transparent). [e] Stability of purified mutant estimated by far-UV CD measurement: U, unstructured; F, folded; PF, partially folded; ND, not determined.

thrombin cleavage and gel filtration. Far-UV CD spectra of the chimeric proteins were measured to estimate their content of ordered structure at 25 °C (Figure 4A). Eight variants of which the second and third segments are derived from POIA1 (XBCXX) show different spectra from the propeptide. Except these XBCXX hybrids, all other 18 hybrids exhibit similar CD spectra, which are typical of largely unstructured proteins such as the propeptide, and they are classified as "U (unstructured)" in Table 1.

Structural properties of XBCXX detected by CD are divided into three groups. The first group consists of ABC45, ABC4E, and ABCD5, which display significant secondary structure to almost the same extent of POIA1 (Figure 4A). The second group, 1BC45 and 1BC4E, show lower ellipticities in absolute value than that of POIA1, suggesting lower content in the secondary structure (Figure 4A). On the other hand, in ^{1}H NMR spectroscopy of 1BC45, the chemical shift dispersion of the resonances in the methyl region observed at −1 to 0.5 ppm shows the characteristics of existence of a tertiary structure (Figure 4B, upper). Among them, it is clearly shown that the location of the β1 strand and α1 helix is identical for 1BC45 and POIA1 since upfield shift of Ile18 of POIA1 (*37*) (Figure 4B, lower) due to shielding of Phe9 stays at −0.9 ppm. These results indicate that 1BC45 and 1BC4E are structured proteins at 25 °C with different contents of secondary structure from POIA1. The first and second groups have the same retention time as POIA1 in gel filtration chromatography, and the ellipticities of the hybrids are independent of concentration between 2 and 100 μM at CD measurement (data not shown), suggesting that they retain compact and monomeric structures just like POIA1. The CD (Figure 4A) and ^{1}H NMR (unpublished data) spectra of the third group, 1BCD5 and 1BCDE, lie between the native and unfolded states. Together with the fact that the retention time in gel filtration of the third group is close to that of the propeptide, it is concluded that the proteins are partially folded. All these data, summarized in Table 1, clearly show that it is necessary for the chimera protein to be folded to serve as an inhibitor, while the unstructured or partially folded protein cannot serve as a long-lived inhibitor.

Among the folded proteins, the turbidity in the skim milk plate for 1BC45, ABC45, and ABCD5 was higher than that for 1BC4E, ABC4E, and POIA1 (Table 1). To seek the origin of the difference in turbidity, we have determined the inhibition constants for the binding of the mutants to subtilisin BPN′ (Table 2). Proteins 1BC45, ABC45, and ABCD5, of which the fifth elements are all derived from the propeptide, have lower K_i(app) values (stronger inhibitory activities) than 1BC4E, ABC4E, and POIA1 by roughly 2 orders of magnitude. On the other hand, thermal denaturation curves for 1BC45 and 1BC4E are almost identical, suggesting that the flexible C-terminal region does not affect the stability of the hybrids (data not shown). Therefore, among structured hybrids, it is indicated that the difference in turbidity comes from the direct interaction between the C-terminal region and the active site of subtilisin (Figure

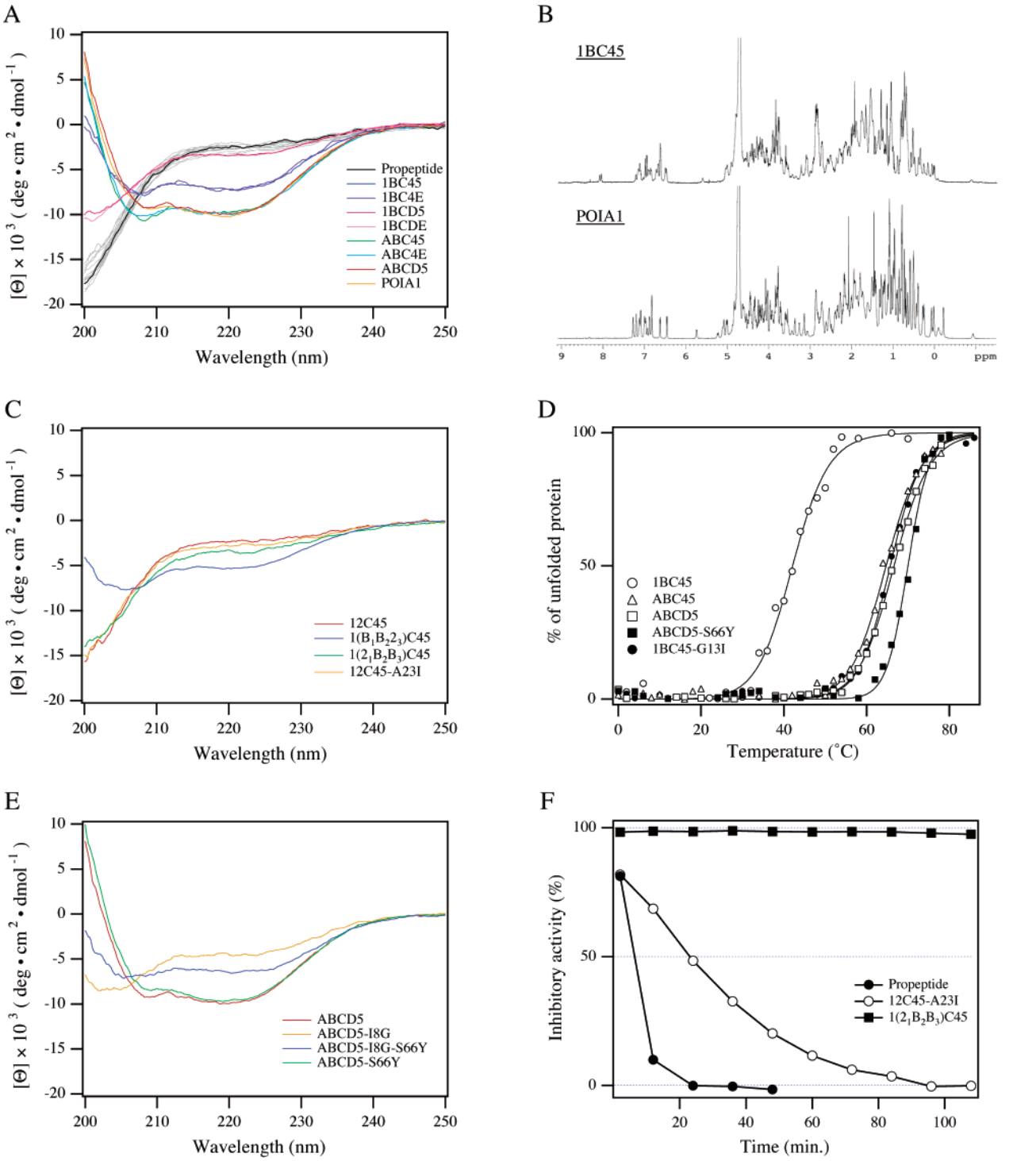

FIGURE 4: Panel A shows CD spectra of hybrids acquired at 25 °C in PBS. Mean residue ellipticity is plotted versus wavelength. Unfolded spectra of the other 15 mutants are depicted in gray. Panel B shows one-dimensional ^{1}H NMR spectra of 1BC45 (upper) and POIA1 (lower) at 25 °C in 99.9% D_2O. Panel C shows CD spectra of additional combinatorial mutants, multisectioned mutants in the second segment ($1(B_1B_22_3)C45$ and $1(2_1B_2B_3)C45$) and a point mutant (12C45-Ala23Ile). Panel D shows thermal unfolding profiles of stable mutants derived from the ellipticity at 220 nm in 50 mM sodium phosphate buffer at pH 7.0. Each solid curve is drawn as a result of curve-fitting based on the van't Hoff equation. T_m values of 1BC45, ABC45, ABCD5, ABCD5-Ser66Tyr, and 1BC45-Gly13Ile are determined to be 42.1, 64.7, 66.6, 70.0, and 65.7 °C, respectively. Panel E shows CD spectra of point mutants of ABCD5. Panel F shows time courses of inhibitory activities of the propeptide, 12C45-Ala23Ile, and $1(2_1B_2B_3)C45$ toward subtilisin BPN′. These three inhibitors have similar apparent inhibitor constants, K_i(app), but show different tolerance to proteolysis. The first measurement was monitored at 2 min after mixing.

1A) and does not reflect the difference in stability of the molecule.

The Most Important Interactions for Foldability. According to the in vitro experiments of purified chimeras, it was shown that the most important interactions for foldability are the interactions between the POIA1-derived second and third segments since XBCXX mutants exclusively showed stability (Table 1). We then tried to specify the inevitable interactions for stability by applying the "multisection" method repeatedly, since it can quickly reach the destination

Table 2: Values of Apparent Inhibition Constants, K_i(app), for the Interaction of Hybrids with Subtilisin BPN′

hybrid protein	K_i(app) (nM)
1BC45	0.060[a]
1BC4E	3.5 ± 0.1
ABC45	0.038[a]
ABC4E	2.2 ± 0.2
ABCD5	0.034[a]
POIA1(ABCDE)	10.6 ± 0.5
propeptide(12345)	0.34 ± 0.03
12C45-Ala23Ile	0.66 ± 0.07
$1(2_1B_2B_3)$C45	0.41 ± 0.04

[a] Results of the average of two experiments calculated with the use of spectrofluorometer.

in an inverse exponential manner. In the present case, additional local segmentation into three fragments in the second element of 1BC45 was introduced (Figure 5A). Two regions identical in the amino acid sequence were utilized as crossover sites (boxes in Figure 5A), and as a result, we constructed six additional variants from $1(2_12_2B_3)$C45 to $1(B_1B_22_3)$C45, in which subscripts represent the partitioned segments in the second element. We then investigated turbidity and foldability for these variants (second row in Table 1). Although three of these mutants ($1(2_1B_22_3)$C45, $1(2_1B_2B_3)$C45, and $1(B_1B_22_3)$C45) show strong endurance against proteolytic degradation in the in vivo selection, only $1(B_1B_22_3)$C45 exhibits a significant fraction of the structured state, while both $1(2_1B_22_3)$C45 and $1(2_1B_2B_3)$C45 can be judged as mostly unstructured according to far-UV CD measurement (Figure 4C). These results indicate that the interaction between the first two-thirds of the second segment (B_1B_2) and third element (C) is necessary for the folding of the protein. This part of the second segment can be divided into two structural components according to the secondary and tertiary structure. One component is the loop (B_1) contacting mainly with Lys42 in the β3 strand in POIA1 (Figure 5B), although this contact appears weak in the propeptide (Figure 5C). The other component is the first half of the α1 helix (B_2), which interacts mainly with Lys42 and Phe44 in the β3 strand in POIA1 (Figure 5B), while in the propeptide it is helical and interacts with Gln40 in the β2 strand (Figure 5C). Existence of these strong interactions was also confirmed by the contact map drawn by MOLMOL (*43*). By combining the results of the multisection method with the insights from the structure, we can thus efficiently specify these important residues in maintaining the structure.

Additional Prominent Sites for Stability. Among 1BCXX mutants, the proteins were foldable only when they have propeptide-derived fourth element (1BC4X) (Table 1), suggesting that a second important interaction for stability exists between the first and the fourth segments. We therefore examined thermostability of XBCX5 except 1BCD5, which is significantly unstable. Temperature unfolding profiles of 1BC45, ABC45, and ABCD5 showed sigmoidal thermal transition curves (Figure 4D) and full reversibility under the condition used. There are obvious differences in stability among the stable mutants. T_m, at which half the native structure is unfolded, of 1BC45, which has a propeptide-derived first segment, is 42.1 °C, while that of ABC45 is 64.7 °C. According to the structural comparison, the first segment has only one secondary structure, β1 strand, and most of the amino acid residues are identical except that Gly13 in the propeptide is replaced with Ile8 in POIA1 (Figure 1D). The β1 strand contacts with the β4 strand in the fourth element, and Gly13 and Tyr67 of the propeptide make a hydrogen bond (Figure 1B), whereas Ile8 and Ser66 of POIA1 form hydrogen bonds (Figure 1C). We therefore introduced point mutations Ile8Gly or Ser66Tyr or both in ABCD5 and examined their stability to identify the combination of these residues necessary for higher stability. Since ABCD5 has Ile8 in the β1 strand and Ser66 in the β4 strand, we termed this combination as "Ile-Ser pair" hereafter. Characterization of ABCD5-Ile8Gly (Gly-Ser pair) by far-UV CD revealed that the protein is a mixture of the native and unfolded structure (Figure 4E), indicating that this one amino acid residue substitution destabilizes ABCD5 significantly to almost the same extent as 1BCD5 (Figure 4A). By applying additional substitution at residue 66 (ABCD5-Ile8Gly-Ser66Tyr), ellipticity was recovered to almost identical extent to 1BC45, which has the same Gly-Tyr pair (Figure 4A,E). It is thus likely that the bulky side chain of tyrosine in the β4 strand can compensate for the instability caused by the loss of the side chain from isoleucine to glycine in the β1 strand. Furthermore, ABCD5-Ser66Tyr, which has the Ile-Tyr pair, melts more than 3 °C higher than ABCD5 (Figure 4D). These results suggest that at least one hydrophobic residue is necessary for overall stability among these pairwise interactions. To validate the legitimacy of this assumption, a point mutant Gly13Ile of 1BC45 (equivalent site to Ile8 in ABCD5) was made. Although 1BC45-Gly13Ile (Ile-Tyr pair) exhibits almost the same far-UV CD spectrum and one-dimensional ^{1}H NMR spectrum as 1BC45 (data not shown), the mutant is stabilized greatly (from 42.1 to 65.7 °C in T_m) to the same extent as ABC45 and ABCD5 (Figure 4D). Therefore it is clear that this isoleucine in the β1 strand is the primary determinant for stability irrespective of the residue in the pair.

Correlation between Foldability and Function. When the multisection method was applied to the second segment of 1BC45, we found that among the singly substituted mutants ($1(2_1B_2B_3)$C45, $1(B_12_2B_3)$C45, and $1(B_1B_22_3)$C45), tolerance to proteolysis by subtilisin BPN′ as judged from the turbidity scale dropped dramatically in the case of $1(B_12_2B_3)$C45 (from ++++ to +) when compared with the other two (to +++ or ++++), indicating that B_2 is essential for maintaining the inhibitory activity (Table 1). Among the B_2 mutants with stronger tolerance, however, only $1(B_1B_22_3)$C45 is partially folded while the other two, $1(2_1B_2B_3)$C45 and $1(2_1B_22_3)$-C45 appear mostly unfolded judging from the far-UV CD measurement (Figure 4C and Table 1). These results indicate that only the mutants that have the POIA1-derived first half of α1 helix (B_2) and third segment (C) show strong tolerance to proteolysis with little relation to self-foldability. We then tried to identify important amino acid residues for resistance toward proteolytic activity by scrutinizing the structure of $1(2_1B_22_3)$C45, which retains inhibitory activity with minimum similarity to POIA1 in the sequence. According to the contact map, in the first half α1 helix (2_2 or B_2) there are two amino acids that contact with the third segment, Ala23 and Lys27 in the propeptide. Among them, Ala23 (Figure 1B) is altered to isoleucine in POIA1 (Figure 1C), while lysine remains the same, and thus, the Ala23Ile substitution has been introduced in 12C45. Although 12C45-Ala23Ile exhibits the characteristics of a disordered structure to the

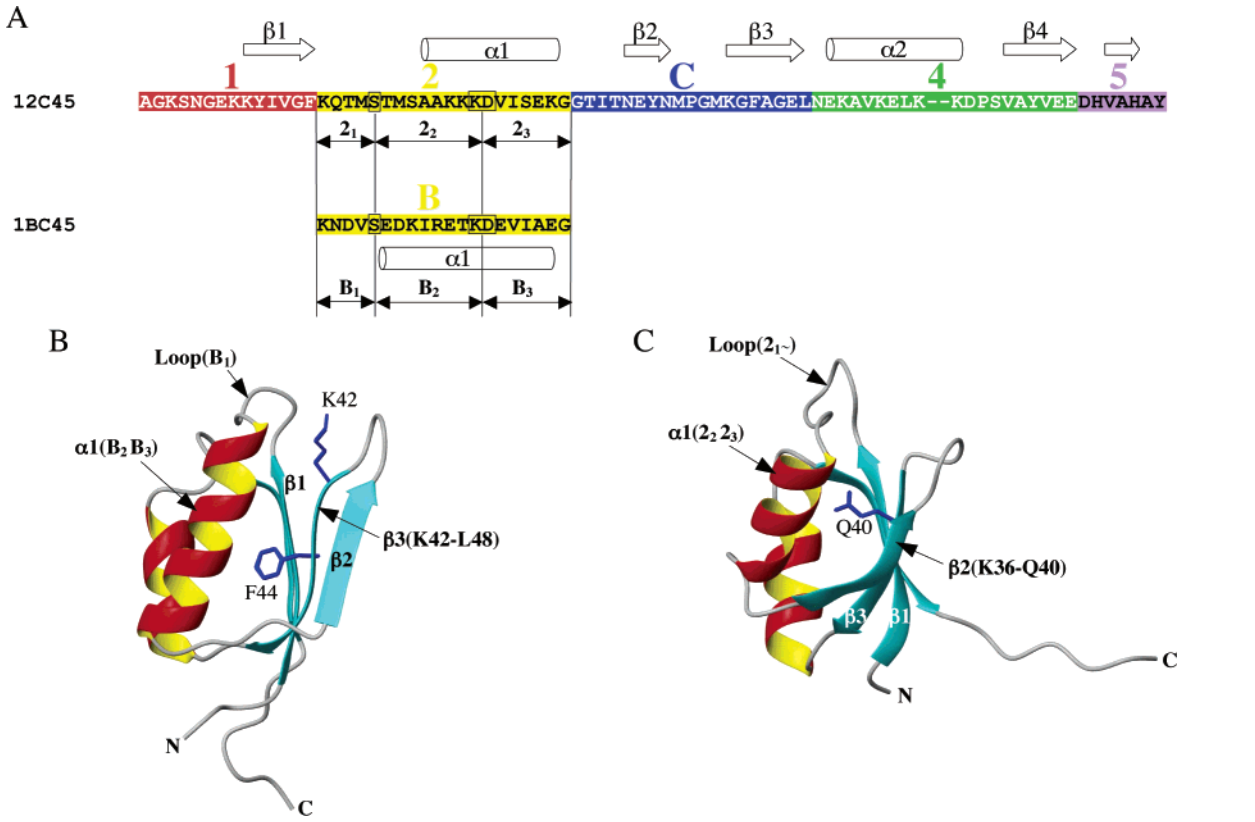

FIGURE 5: Panel A illustrates the "multisection" method: additional local segmentation in the second segment (2 or B) into three fragments (1−3 in subscripts) was introduced in 1BC45. Amino acid residues enclosed in boxes indicate the boundaries used for this partitioning. These fragments were recombined to produce 8 (= 2^3) variants including 12C45 and 1BC45. Panel B shows the structure of POIA1 viewed from the α1 helix side. Side chains of Lys42 and Phe44, which strongly contact with the loop (B_1) and the first half of the α1 helix (B_2), are shown in blue. Panel C shows the corresponding structure of the propeptide. Gln40, which interacts with the first half of the α1 helix (2_2), is shown in blue.

same extent as the propeptide and 12C45 (Figure 4C), this mutant is able to sustain inhibitory activity longer than the propeptide according to the in vivo selection (++ as shown in the third row in Table 1), suggesting that Ala23Ile in this unstructured protein plays important role to gain tolerance to proteolysis. We then examined kinetic profiles in the course of proteolysis of 12C45-Ala23Ile, 1($2_1B_2B_3$)C45, and the propeptide to know the cause for the apparent contradiction between the tolerance and foldability. It was shown that both 12C45-Ala23Ile and 1($2_1B_2B_3$)C45, especially the latter, have more resistance toward proteolytic activity than the propeptide (Figure 4F), although they bind to subtilisin more weakly than the propeptide (Table 2). It is thus clear that the difference in tolerance shown by turbidity in the in vivo screening among these mutants comes from the difference in degree of endurance against proteolysis. Contrary to the report that the mutants with increased structural content within the propeptide enhance its resistance toward proteolytic degradation (*31*, *34*), these results of 1($2_1B_2B_3$)C45, 1($2_1B_2 2_3$)C45, and 12C45-Ala23Ile show the first example that the tolerance to proteolysis is not necessarily correlated with the independent folding of the molecule.

Combining all results, we propose a plausible mechanism that can correlate the self-foldability with the function (Figure 6). In the overall maturation process of subtilisin, the propeptide acts as an inhibitor after the autoprocessing, and then it is degraded by the mature enzyme. For higher inhibitory activity, it is important that the protein has self-foldability, as well as the propeptide-derived C-terminal

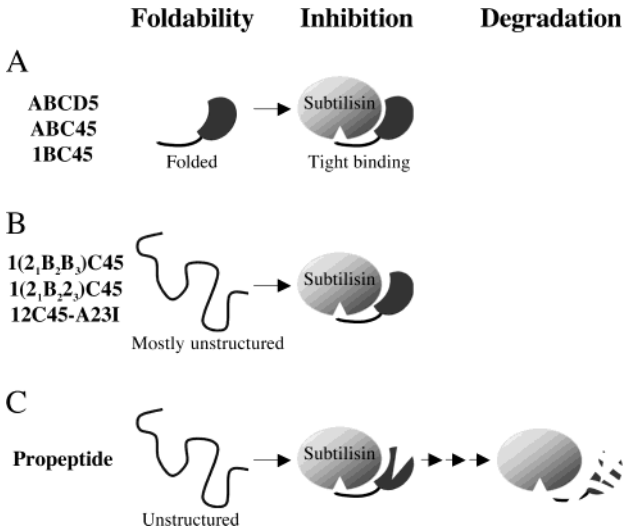

FIGURE 6: Schematic representation of functional roles of the propeptide and its mutants. In panel A, ABCD5, ABC45, and 1BC45 show self-foldability and high affinities for subtilisin. Panel B shows 1($2_1B_2B_3$)C45 and 1($2_1B_2 2_3$)C45, which are unstructured but resistant against proteolysis. 12C45-Ala23Ile is also unstructured with relatively shorter lifetime as an inhibitor (Figure 4F). The hydrophobic interaction in interior of mutant when complexed with subtilisin rather than self-foldability provides its resistance toward proteolytic degradation. Panel C shows the propeptide. Cleft in the propeptide formed upon complex formation shows lack of interaction between the α1 helix and *β*2 and *β*3 strands.

region to interact directly with the active site of the enzyme (ABCD5, ABC45, and 1BC45 in Figure 6A). The unstruc-

tured proteins have weaker binding affinities since they need extra free energy of folding upon binding to the target. To make the propeptide foldable, interactions between the loop-α1 and β2-turn-β3 and between the pair of amino acid residues in the β1 and β4 strands must overcome dynamic fluctuation induced by the exposure of the hydrophobic amino acid residues on the surface of the β sheet that are largely buried upon binding with subtilisin (Figure 1A). Then, why does the propeptide not have self-foldability? We consider that the propeptide has chosen the strategy for higher vulnerability to proceed with the subsequent autodegradation smoothly, rather than higher stability to attain tight binding. Among the unstructured proteins that appear to be vulnerable by the enzyme attack, we have found stronger hydrophobic interaction between the α1 helix and β2 and β3 strands formed upon complex formation prevents autodegradation ($1(2_1B_2B_3)$C45, $1(2_1B_22_3)$C45, and 12C45-Ala23Ile in Figure 6B), suggesting that subtle balance between stability and instability controls total function as a propeptide (Figure 6C). Overall, examination of the hybrids successfully identified necessary sites for not only self-foldability but also function, which is correlated with the "instability" of the protein.

REFERENCES

1. Zhao, H., and Arnold, F. H. (1997) Combinatorial protein design: strategies for screening protein libraries, *Curr. Opin. Struct. Biol. 7*, 480−485.
2. Giver, L., and Arnold, F. H. (1998) Combinatorial protein design by in vitro recombination, *Curr. Opin. Chem. Biol. 2*, 335−338.
3. Hoess, R. H. (2001) Protein design and phage display, *Chem. Rev. 101*, 3205−3218.
4. Moffet, D. A., and Hecht, M. H. (2001) De novo proteins from combinatorial libraries, *Chem. Rev. 101*, 3191−3203.
5. Stemmer, W. P. (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution, *Proc. Natl. Acad. Sci. U.S.A. 91*, 10747−10751.
6. Zhao, H., Giver, L., Shao, Z., Affholter, J. A., and Arnold, F. H. (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination, *Nat. Biotechnol. 16*, 258−261.
7. Crameri, A., Raillard, S. A., Bermudez, E., and Stemmer, W. P. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution, *Nature 391*, 288−291.
8. Ostermeier, M., Shim, J. H., and Benkovic, S. J. (1999) A combinatorial approach to hybrid enzymes independent of DNA homology, *Nat. Biotechnol. 17*, 1205−1209.
9. Lutz, S., Ostermeier, M., Moore, G. L., Maranas, C. D., and Benkovic, S. J. (2001) Creating multiple-crossover DNA libraries independent of sequence identity, *Proc. Natl. Acad. Sci. U.S.A. 98*, 11248−11253.
10. Sieber, V., Martinez, C. A., and Arnold, F. H. (2001) Libraries of hybrid proteins from distantly related sequences, *Nat. Biotechnol. 19*, 456−460.
11. Bogarad, L. D., and Deem, M. W. (1999) A hierarchical approach to protein molecular evolution, *Proc. Natl. Acad. Sci. U.S.A. 96*, 2591−2595.
12. Voigt, C. A., Martinez, C., Wang, Z. G., Mayo, S. L., and Arnold, F. H. (2002) Protein building blocks preserved by recombination, *Nat. Struct. Biol. 9*, 553−558.
13. O'Maille, P. E., Bakhtina, M., and Tsai, M. D. (2002) Structure-based combinatorial protein engineering (SCOPE), *J. Mol. Biol. 321*, 677−691.
14. Hiraga, K., and Arnold, F. H. (2003) General method for sequence-independent site-directed chimeragenesis, *J. Mol. Biol. 330*, 287−296.
15. Wells, J. A., Ferrari, E., Henner, D. J., Estell, D. A., and Chen, E. Y. (1983) Cloning, sequencing, and secretion of *Bacillus amyloliquefaciens* subtilisin in *Bacillus subtilis*, *Nucleic Acids Res. 11*, 7911−7925.
16. Vasantha, N., Thompson, L. D., Rhodes, C., Banner, C., Nagle, J., and Filpula, D. (1984) Genes for alkaline protease and neutral protease from *Bacillus amyloliquefaciens* contain a large open reading frame between the regions coding for signal sequence and mature protein, *J. Bacteriol. 159*, 811−819.
17. Power, S. D., Adams, R. M., and Wells, J. A. (1986) Secretion and autoproteolytic maturation of subtilisin, *Proc. Natl. Acad. Sci. U.S.A. 83*, 3096−3100.
18. Ikemura, H., Takagi, H., and Inouye, M. (1987) Requirement of pro-sequence for the production of active subtilisin E in *Escherichia coli*, *J. Biol. Chem. 262*, 7859−7864.
19. Ikemura, H., and Inouye, M. (1988) In vitro processing of pro-subtilisin produced in *Escherichia coli*, *J. Biol. Chem. 263*, 12959−12963.
20. Zhu, X. L., Ohta, Y., Jordan, F., and Inouye, M. (1989) Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process, *Nature 339*, 483−484.
21. Strausberg, S., Alexander, P., Wang, L., Schwarz, F., and Bryan, P. (1993) Catalysis of a protein folding reaction: thermodynamic and kinetic analysis of subtilisin BPN′ interactions with its propeptide fragment, *Biochemistry 32*, 8112−8119.
22. Ohta, Y., Hojo, H., Aimoto, S., Kobayashi, T., Zhu, X., Jordan, F., and Inouye, M. (1991) Pro-peptide as an intramolecular chaperone: renaturation of denatured subtilisin E with a synthetic pro-peptide, *Mol. Microbiol. 5*, 1507−1510.
23. Wang, L., Ruvinov, S., Strausberg, S., Gallagher, D. T., Gilliland, G., and Bryan, P. N. (1995) Prodomain mutations at the subtilisin interface: correlation of binding energy and the rate of catalyzed folding, *Biochemistry 34*, 15415−15420.
24. Kojima, S., Minagawa, T., and Miura, K. (1997) The propeptide of subtilisin BPN′ as a temporary inhibitor and effect of an amino acid replacement on its inhibitory activity, *FEBS Lett. 411*, 128−132.
25. Fu, X., Inouye, M., and Shinde, U. (2000) Folding pathway mediated by an intramolecular chaperone. The inhibitory and chaperone functions of the subtilisin propeptide are not obligatorily linked, *J. Biol. Chem. 275*, 16871−16878.
26. Yabuta, Y., Takagi, H., Inouye, M., and Shinde, U. (2001) Folding pathway mediated by an intramolecular chaperone: propeptide release modulates activation precision of pro-subtilisin, *J. Biol. Chem. 276*, 44427−44434.
27. Gallagher, T., Gilliland, G., Wang, L., and Bryan, P. (1995) The prosegment-subtilisin BPN′ complex: crystal structure of a specific 'foldase', *Structure 3*, 907−914.
28. Bryan, P., Wang, L., Hoskins, J., Ruvinov, S., Strausberg, S., Alexander, P., Almog, O., Gilliland, G., and Gallagher, T. (1995) Catalysis of a protein folding reaction: mechanistic implications of the 2.0 Å structure of the subtilisin-prodomain complex, *Biochemistry 34*, 10310−10318.
29. Eder, J., Rheinnecker, M., and Fersht, A. R. (1993) Folding of subtilisin BPN′: role of the pro-sequence, *J. Mol. Biol. 233*, 293−304.
30. Ruan, B., Hoskins, J., and Bryan, P. N. (1999) Rapid folding of calcium-free subtilisin by a stabilized pro-domain mutant, *Biochemistry 38*, 8562−8571.
31. Ruvinov, S., Wang, L., Ruan, B., Almog, O., Gilliland, G. L., Eisenstein, E., and Bryan, P. N. (1997) Engineering the independent folding of the subtilisin BPN′ prodomain: analysis of two-state folding versus protein stability, *Biochemistry 36*, 10414−10421.
32. Wang, L., Ruan, B., Ruvinov, S., and Bryan, P. N. (1998) Engineering the independent folding of the subtilisin BPN′ pro-domain: correlation of pro-domain stability with the rate of subtilisin folding, *Biochemistry 37*, 3165−3171.
33. Ruan, B., Hoskins, J., Wang, L., and Bryan, P. N. (1998) Stabilizing the subtilisin BPN′ pro-domain by phage display selection: how restrictive is the amino acid code for maximum protein stability?, *Protein Sci. 7*, 2345−2353.
34. Kojima, S., Minagawa, T., and Miura, K. (1998) Tertiary structure formation in the propeptide of subtilisin BPN′ by successive amino acid replacements and its close relation to function, *J. Mol. Biol. 277*, 1007−1013.
35. Dohmae, N., Takio, K., Tsumuraya, Y., and Hashimoto, Y. (1995) The complete amino acid sequences of two serine proteinase inhibitors from the fruiting bodies of a basidiomycete, *Pleurotus ostreatus*, *Arch. Biochem. Biophys. 316*, 498−506.
36. Dohmae, N., Hayashi, K., Miki, K., Tsumuraya, Y., and Hashimoto, Y. (1995) Purification and characterization of intracellular proteinases in *Pleurotus ostreatus* fruiting bodies, *Biosci., Biotechnol., Biochem. 59*, 2074−2080.
37. Sasakawa, H., Yoshinaga, S., Kojima, S., and Tamura, A. (2002) Structure of POIA1, a homologous protein to the propeptide of

subtilisin: implication for protein foldability and the function as an intramolecular chaperone, *J. Mol. Biol. 317*, 159−167.
38. Lo Conte, L., Brenner, S. E., Hubbard, T. J., Chothia, C., and Murzin, A. G. (2002) SCOP database in 2002: refinements accommodate structural genomics, *Nucleic Acids Res. 30*, 264−267.
39. Gill, S. C., and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data, *Anal. Biochem. 182*, 319−326.
40. Taguchi, S., Ozaki, A., and Momose, H. (1998) Engineering of a cold-adapted protease by sequential random mutagenesis and a screening system, *Appl. Environ. Microbiol. 64*, 492−495.
41. Henderson, P. J. (1972) A linear equation that describes the steady-state kinetics of enzymes and subcellular particles interacting with tightly bound inhibitors, *Biochem. J. 127*, 321−333.
42. Kojima, S., Hisano, Y., and Miura, K. (2001) Alteration of inhibitory properties of *Pleurotus ostreatus* proteinase A inhibitor 1 by mutation of its C-terminal region, *Biochem. Biophys. Res. Commun. 281*, 1271−1276.
43. Koradi, R., Billeter, M., and Wuthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures, *J. Mol. Graphics 14*, 51−55.

BI0302402