J. Biochem. 128, 745–754 (2000)

Functional Mapping against *Escherichia coli* for the Broad-Spectrum Antimicrobial Peptide, Thanatin, Based on an *In Vivo* Monitoring Assay System[1]

Seiichi Taguchi,*[,2] Kanako Kuwasako,* Atsushi Suenaga,† Miyuki Okada,* and Haruo Momose*

Department of Biological Science and Technology, Science University of Tokyo, 2641 Yamazaki, Noda, Chiba 278-8510; and †Computational Science Division, Advanced Computing Center, RIKEN Institute, 2-1 Hirosawa, Wako, Saitama 351-0198

Received July 7, 2000; accepted August 18, 2000

Previously, we established for the first time an *in vivo* monitoring assay system conjugated with random mutagenesis in order to study the structure–function relationship of the antimicrobial peptide, apidaecin [Taguchi *et al.* (1996) *Appl. Environ. Microbiol.* 62, 4652–4655]. In the present study, this methodology was used to carry out the functional mapping of a second target, thanatin, a 21-residue peptide that exhibits the broadest antimicrobial spectrum so far observed among insect defense peptides [Fehlbaum *et al.* (1996) *Proc. Natl. Acad. Sci. USA* 93, 1221–1225]. First, a synthetic gene encoding thanatin was expressed in a fused form with *Streptomyces* protease inhibitor protein, SSI, under the control of *tac* promoter in *Escherichia coli* JM109. Expression of the thanatin-fused protein was found to depend on the concentration of the transcriptional inducer, isopropyl-β-D-thio-galactopyranoside (IPTG), and to parallel the degree of growth inhibition of the transformant cells. When a PCR random mutation was introduced into the structural gene for thanatin, diminished growth inhibition of the IPTG-induced transformed cells was mostly observed in variants as measured by colony size (plate assay) or optical density (liquid assay) in comparison with the wild-type peptide, possibly depending on the decreased antimicrobial activity of each variant. Next, wild-type thanatin and three variants screened by the *in vivo* assay, two singly mutated proteins (C11Y and M21R) and one doubly mutated protein (K17R/R20G), were stably overproduced with a fusion partner protein resulting in the efficient formation of inclusion bodies in *E. coli* BL21(DE3). The products were isolated in large amounts (yield 30%) from the fused protein by successive chemical and enzymatic digestions at the protein fusion linker site. Anti-*E. coli* JM109 activities, judged by minimum inhibitory concentration, of the purified peptides were in good agreement with those estimated semi-quantitatively by the *in vivo* assay. Based on the NMR solution structure and molecular dynamics, the structure–function relationship of thanatin is discussed by comparing the functional mapping data obtained here with the previous biochemical data. The functional mapping newly suggests the importance of a hydrogen bonding network formed within the C-terminal loop joining the β-strands arranged antiparallel to one another that are supposed to be crutial for exhibiting anti-*E. coli* activity.

Key words: antimicrobial peptide, molecular dynamics simulation, NMR solution structure, PCR random mutagenesis, structure–function relationship, thanatin.

Insects have been remarkably successful in evolution. Their host-defense systems rely on several innate reactions upon injury by an extremely large variety of potentially harmful microorganisms. Rapid and transient synthesis of secretory antimicrobial peptides is a very effective defense strategy exhibiting a broad spectrum of activity directed against bacteria and/or fungi (*1*).

For the study of antimicrobial peptides, chemical synthesis is a very useful approach for establishing the structure and composition of natural products, for making hybrid or chimeric peptides, for defining details of precursor processings, and for studies on the mechanisms of action of the peptides. However, this approach has some limiting factors, for instance, the high cost and difficulties in designing variants at the level of single amino acid without detailed information concerning the structure–function relationship or molecular phylogenetic data. We have already developed an *in vivo* system for monitoring antimicrobial activity tightly

[1] This study was supported in part (to S.T.) by a Grant-in-Aid for Scientific Research (no. 70216828) from the Ministry of Education, Science, Sports and Culture of Japan, and a grant from the Nissan Science Foundation (Tokyo).

[2] To whom correspondence should be addressed at the present address: Polymer Chemistry Laboratory, RIKEN Institute, 2-1 Hirosawa, Wako, Saitama 351-0198. Tel: +81-48-467-9404, Fax: +81-48-462-4667, E-mail: staguchi@postman.riken.go.jp

correlated to changes in the growth inhibition (suicide effect by antimicrobial action) of host cells expressing wild-type or variants of apidaecin (*2*, *3*). This system is characterized by the use of *Escherichia coli* which are intrinsically sensitive to apidaecin, as the recipient host. This is a suicide system carrying a toxic molecule whose production can be stringently controlled by transcriptional regulation.

In this study, we selected a second target, thanatin, as a model antimicrobial peptide to elucidate the versatility of our *in vivo* functional mapping methodology. Thanatin is a 21-residue inducible defense peptide found in the hemipteran insect *Podisus maculiventris*. Thanatin exhibits the broadest range of antimicrobial activity (bactericidal and fungicidal) so far characterized (*4*). Biochemical and structural studies of thanatin were carried out by synthesizing truncated isoforms (*4*) and by using two-dimentional (2D) ^{1}H-NMR and a molecular modeling method (*5*). Here we describe the recombinant overproduction, purification and characterization of wild-type thanatin together with its variants screened through the *in vivo* functional mapping approach. The structure–function relationship of thanatin is deduced by comparing the data with those reported previously (*4*, *5*).

MATERIALS AND METHODS

Materials—All restriction enzymes and modifying enzymes for genetic engineering were purchased from TaKaRa Shuzo (Kyoto). Crystallographic grade proline iminopeptidase was kindly supplied by Dr. T. Yoshimoto, Nagasaki University. The polyclonal antibody against *Streptomyces* subtilisin inhibitor (SSI) was raised in rabbits and its reactivity with the antigen was checked in our laboratory. All other chemicals were of analytical grade for biochemical use and were used without further purification.

Expression Systems—The preparation of plasmid DNA from *Escherichia coli* and the transformation of *E. coli* were carried out according to standard procedures (*6*). The four individual single-stranded oligomers encoding thanatin with the 70% formic acid selective cleavage site (Asp-Pro) and two restriction cleavage sites (*Eco*RI and *Bam*HI) to be assembled (Fig. 1), were synthesized by the solid phase phosphoamidite method with an Applied Biosystems 381A DNA synthesizer. The resulting DNA was cloned into pUC18 using the *Eco*RI/*Bam*HI sites, and the resultant was termed pU-tan. To construct an *in vivo* monitoring assay system for thanatin, two plasmid vectors, pOSBΔ-AP1 (*2*, *3*) and pMKSI161-9 (*7*), were used as illustrated in Fig. 1. The former vector allows the functional expression for antimicrobial peptide, apidaecin, by efficient control of *lac* promoter-operator and OmpA signal sequence (*8*). The latter is a secretory expression vector for the stable protease inhibitor, SSI, using *tac* promoter and *lac* operator. Both vectors allow gene expression that can be controlled by induction with isopropyl-1-thio-β-D-galactopyranoside (IPTG). These vectors were prepared to achieve the production in *E. coli* of small-sized peptides of interest as fused forms with SSI. The former vector is an improved version of the latter in the production level of SSI (roughly estimated to be 200 to 1). The thanatin encoding region on pU-tan was transferred to the same *Eco*RI/*Bam*HI sites of pOSΔB-AP1 and pMKSI161-9 to generate pOS-tan and pMKSI-tan, respectively. For cultivation, minimum medium supplemented with casamino acid (DM-CA medium) was used to perform the stringent expression of the thanatin gene without the leaking at the transcription level that often occurs due to an unkown inducer-like compound in rich media such as LB. DM-CA medium (adjusted to pH 7.0) comprises 0.7% Na_2HPO_4, 0.2% KH_2PO_4, 0.05% sodium citrate–$2H_2O$, 0.1% $(NH_4)_2SO_4$, 0.01% $MgSO_4$–$7H_2O$, 0.0005% thiamine-HCl, and 0.2% casamino acid. The expression and localization (cytoplasm, periplasm, extracellular culture medium) of the SSI-tanantin (tan) fused protein were checked by electrophoretic and immunological analyses as described previously (*2*, *3*). Transformants of *E. coli* JM109 producing thanatin variants with varying activities were screened for colony size or optical density related to the degree of antimicrobial activity of each thanatin variant on agar plates (primary screening) or in liquid medium (secondary screening), respectively. Both screenings were conducted under conditions similar to those described previously (*2*, *3*).

In Vitro Random Mutagenesis—The plasmid vector containing the thanatin gene, pU-tan, was used as a template for PCR mutagenesis with two universal sequencing primers (TaKaRa Shuzo), M4 (near the *Bam*HI site) and RV (near the *Eco*RI site). The target region, including the thanatin gene, was amplified with the two primers under error-prone conditions: addition of 1% β-mercaptoethanol, 0.25 mM $MnCl_2$, and 10% dimethyl sulfoxide to the 100 μl reaction buffer containing 2.5 nM pU-tan, 0.025 U of *Taq* polymerase, 1 μM primers, 0.2 mM each deoxynucleotide triphosphate, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, and 1.5 mM $MgCl_2$. PCR was carried out using a program of 35 cycles of 93°C for 90 s, 43°C for 30 s, and 72°C for 3 min with a Program Temp Control System PC-700 (ASTEC Co.). The mutation points were analyzed by dideoxynucleotide chain–termination sequencing using a BcaBEST kit (TaKaRa Shuzo). A list of mutations is presented in Fig. 2. One single-mutation–carrying thanatin, M21R (abbreviation for amino acid substitution at position 21, Met21 to Arg) was newly generated by separation from the triple mutant, S2P/Y10F/M21R, using a unique *Spl*I restriction site in the thanatin gene (Fig. 1). *Spl*I is useful for checking the presence and mutation of the thanatin gene during subcloning and PCR mutagenesis. Thanatin variants were produced by recombination for the purpose of *in vivo* (small scale) and *in vitro* (large scale) assays by the same procedures employed for the wild-type thanatin.

Peptide Purification and Isolation—For biochemical characterization of the thanatin peptide, the genes encoding wild-type and variants were inserted into the plasmid vector, pMTAKHV1 (Fig. 1). Overproduction of wild-type and variant thanatins was achieved in *E. coli* BL21(DE3) using a high-level expression plasmid vector, pMTAKHV1 (*9*), leading to the formation of inclusion bodies. This vector was developed for the high-level production of a human thrombin inhibitor protein, hirudin, as a fused form with porcine muscle adenylate kinase. The resultant plasmids, pAHS-tan and its derivatives, were introduced into *E. coli* BL21-(DE3). A transformant harboring pAHS-tan or a derivative was cultivated in 100 ml of M9 medium (*6*) containing 0.2% glucose, 0.2% casamino acid, 1 mM thiamine-HCl, 1 mM $MgSO_4$–$7H_2O$, 0.1 mM $CaCl_2$, and ampicillin (100 μg/ml) in each batch (total 10 Sakaguchi's flasks) at 37°C. Induction with indoleacetic acid (IAA), a tryptophan derivative, was

done at a final concentration of 0.05 mM when optical density at 610 nm of the culture reached 0.5. This was followed by additional cultivation for 5 h at 37°C. Cells were collected by centrifugation, sonicated and centrifuged to obtain the insoluble fraction containing inclusion bodies, as described previously (*9*). The aimed fraction of the fused-protein was checked by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and Western blotting. The insoluble fraction was washed twice with 5 ml of buffer solution [10 mM KH_2PO_4 (pH 7.0), 1 mM EDTA, 4% Triton X-100]. The Triton X-100 was removed with 25 ml of distilled water, and the fraction was subjected to 2,500-×*g* centrifugation. The protein precipitate was solubilized in 30 ml of 8 M urea solution containing 0.1 M β-mercaptoethanol and 50 mM Tris-HCl (pH 8.5), and concentrated by 10% trichloroacetic acid (TCA) precipitation. The fused protein sample was dissolved in 70% formic acid and a chemical reaction was done to isolate the Pro-joined thanatin peptide at 37°C for about 72 h, referring to the previous report (*10*). The recovery of Pro-joined thanatin was monitored by SDS-PAGE and liquid bioassay. Complete purification of the Pro-joined form was carried out by reversed-phase high-performance liquid chromatography (RP-HPLC) on a Nakarai C_{18} column (250 × 4.6 mm; Nakaraitesk) at a flow rate of 0.7 ml/min using adaptive linear gradients of acetonitrile in 0.1% trifluoroacetic acid (TFA) over 85 min: (i) 0 to 3 min, 0%; (ii) 3 to 15 min, 0 to 16%; (iii) 15 to 75 min, 16 to 24%; (iv) 75 to 80 min, 24 to 80%; (v) 80 to 85 min, 80 to 0%. UV absorption at 225 nm and antimicrobial activity were routinely monitored. The recovered Pro-joined thanatin sample was evaporated once and subjected to enzymatic digestion with proline iminopeptidase. Removal of the Pro residue was done in 300 μl of Tris-HCl (pH 8.0) buffer solution containing 4,700 pmol of peptide by the addition of 4 U of proline iminopeptidase (*11*) at 37°C overnight. The enzymatic digest was again applied to RP-HPLC performed under the same condition as above to isolate the native thanatin. As for peptide concentration, the absorbance at 214 nm for 10 μg of pure thanatin is 192 mOD using a narrow column (2.1 mm of internal diameter) with an acetonitrile gradient from 2 to 19% in 10 min plus an additional step from 19 to 29% in 50 min. When the column work is done at 225 nm under the same conditions, mOD of 98 corresponds to 10 μg. The precise determination of the peptide concentration of thanatin derivatives can also be done by amino acid analysis using an protein sequence analyzer.

Bioassay—Antimicrobial activity was monitored during the purification process by a liquid-growth inhibition assay using *E. coli* JM109 according to the previous procedure (*4*) with slight modification. Briefly, the assay was performed by adding 20 μl of prefiltered peptide sample, at various concentrations, to 80 μl of PB medium (1% Bacto tryptone and 0.9% NaCl) including 300 cells of the test strain, followed by 12-h incubation at 25°C with gentle shaking in a 96-well micro-titer plate. To check for bacterial contamination in the peptide samples, PB medium alone was used as a negative control. Experimental reproducibility was confirmed by two trials.

Analytical Procedures—SDS-PAGE was carried out according to the method of Laemmli (*12*) with a polyacrylamide concentration of 18.8%. The protein samples were precipitated with TCA at a final concentration of 8%. Proteins in the gel were stained with 2% (w/v) Coomassie Brilliant Blue R250. The prepared anti-SSI polyclonal antibody was used for Western blot analysis of thanatin-fused protein with SSI as described previously (*7*). The amino acid sequences of the purified thanatin and its variants were analyzed using a protein sequencer (model 477A, Perkin-Elmer, USA). Sequencing was carried out according to the sequencer manual, and phenylthiohydantoin (PTH)–amino acids were identified with an on-line HPLC system (model 120A, Perkin-Elmer). Amino acid analysis was performed with an amino acid analyzer model A8700 (Irica Instrument, Kyoto) at a sensitivity setting of 2.5 nmol full scale. The analyzer was equipped with a cation-exchange chromatographic column and a standard ninhydrin detection system. Samples for analysis were hydrolyzed *in vacuo* at 110°C for 24 h in 6 N HCl. Fast atom bombardment mass spectrometry (FAB-MS) was performed on an HX-110 mass spectrometer (JEOL, Tokyo), equipped with a DA5000 data system, employing an accelerating voltage of 10 kV and xenon as an ionizing gas. The reaction mixture was dissolved in 2 μl of 67% acetic acid and mixed with 1 μl of matrix consisting of glycerol, thioglycerol, and *m*-nitrobenzylalcohol (1:1:1, by volume).

Molecular Dynamics Simulation—The NMR solution structure of thanatin [PDB code; 8tfv (*5*)] was used as the initial structure for molecular dynamics (MD) simulation. To prepare the C11Y and K17R/R20G variants of thanatin, Cys11 was replaced by Tyr, and Lys17/Arg20 were replaced by Arg and Gly, respectively. To neutralize the systems, counter ions (Na^+ or Cl^-) were placed near the charged residues. The initial structures including variants (C11Y and K17R/R20G) were placed in boxes filled with a TIP3P water model (*13*). The size of the boxes, 56 × 39 × 36, was chosen so that the distance of the atoms in the protein from the wall was greater than 10.0 Å. The total number of atoms for the simulations was as large as approximately 7,500. We adopted the force field of Cornell *et al.* (*14*), and imposed the periodic boundary condition. All calculations were performed on a Fujitsu VPP700E supercomputer with an Amber 5.0 simulation package (*15*). The temperature and pressure were kept constant (at 300 K and 1 atm, respectively) by the method of Berendsen *et al.* (*16*), and the solute and solvent were coupled separately to a temperature bath with coupling constants of 0.2 ps and a pressure relaxation time of 1.9 ps in all simulations. The particle-mesh Ewald was used to calculate the long-range Coulomb force. Only bonds involving hydrogen atoms were constrained by the SHAKE method (*17*), and the time step was set at 0.5 fs.

RESULTS AND DISCUSSION

Randomly Mutagenized Variants of Thanatin Screened by an In Vivo Monitoring Assay System—In the initial step, for the construction of *in vivo* assay system, we attempted to achieve the functional expression of a synthetic thanatin gene as a fused form with the gene for a stable partner protein SSI using the vector pOS-tan (Fig. 1), a derivative of pOSΔB-AP1 that had given the successful secretory expression of an antimicrobial peptide, apidaecin (*2*, *3*). However, no transformant cells harboring pOS-tan appeared on a selective plate. This suggested that even slight leakiness of the expression of a thanatin gene product with a strong bactericidal activity would lead to the death of the host

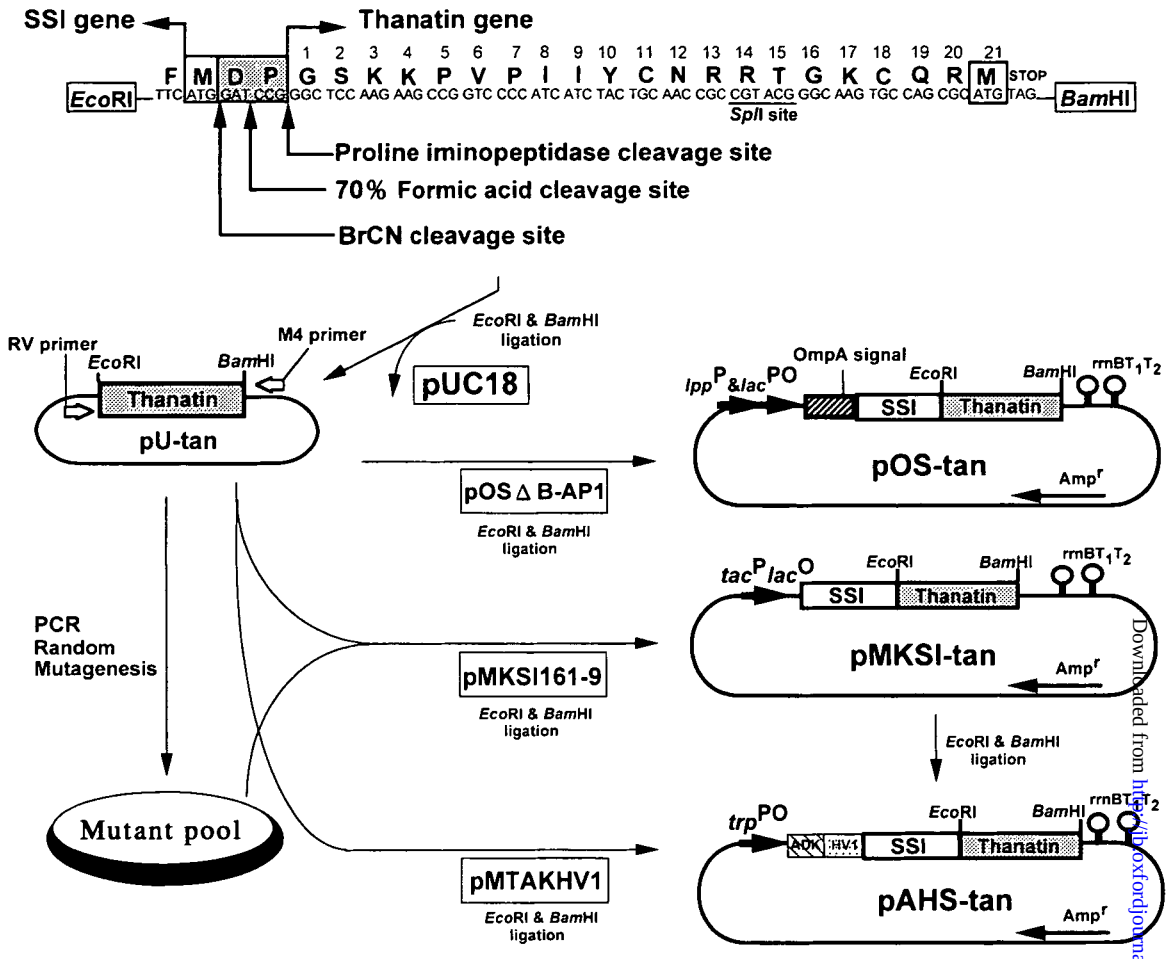


Fig. 1. **Expression vector constructs.** Primary structure of the chemically synthesized thanatin gene shown with restriction enzyme sites *Eco*RI and *Bam*HI. The thanatin gene contains a unique *Spl*I site. A full explanation of the construction of three expression plasmid vectors of thanatin-fused proteins, pOSΔB-tan (for *in vivo* assay), pMKSI-tan (for *in vivo* assay) and pAHS-tan (for overproduction and *in vitro* assay), is given in "MATERIALS AND METHODS." A mixture of variants with heterogeneous mutations was introduced into the pMKSI-161-9 to generate derivatives of pMKSI-tan.

cells, as compared to apidaecin, which has a weak bactriostatic activity (*2*, *3*). To prevent the lethal expression of the thanatin gene, the synthetic gene was displaced in frame to the end of the structural gene for SSI (SSI-tan) on a low-level expression plasmid vector, pMKSI161-9 (*7*), to generate pMKSI-tan, as shown in Fig. 1. This alteration allowed us to carry out functional mapping in terms of colony size, that is, growth inhibition was clearly observed only when the transcriptional inducer IPTG for the SSI-tan fused gene was added to the transformant cells harboring pMKSI-tan. The degree of growth inhibition was, as expected, dependent on the concentration of IPTG (data not shown). On the other hand, the control strain producing SSI alone showed almost no growth inhibition. Therefore, in establishing the *in vivo* assay system, attention should be paid to fine-tuning the expression level in the host strain tested of the antimicrobial peptide of interest according to its activity.

We next introduced *in vitro* random mutations *via* an error-prone PCR into the *Eco*RI–*Bam*HI DNA fragment containing the entire thanatin gene on pU-tan (Fig. 1). The mutagenized DNA fragments, including a mixture of mutated thanatin genes, were ligated into pMKSI161-9 using the *Eco*RI and *Bam*HI sites, and then introduced into *E. coli* JM109. The resulting ampicillin-resistant transformant colonies formed on the plates containing IPTG (at a rather stringent concentration of 0.05 mM). Through PCR random mutagenesis, thirty candidates with possibly reduced activities were obtained from primary screening based on the plate assay system. The frequency of the appearance of colonies on plates containing 0.05 mM IPTG relative to that on plates without IPTG was calculated to be approximately 0.2%. Nucleotide sequence analysis revealed that 16 single-mutation–carrying variants, 3 double-mutation–carrying variants, and 3 triple-mutation–carrying variants were included, as summarized in Fig. 2. Of the 16 variants with single mutations, 4 had in multiple the identical single mutations as follows, I9N: 4, R13H: 2, G16C: 3, and G16V: 3.

Figure 3 shows the growth curves of transformant cells harboring plasmid vectors carrying the wild-type thanatin gene and 23 mutated variants, including a genetically generated M21R, together with pMKSI161-9 (SSI gene alone) as a control, grown in a liquid medium. The growth inhibition patterns observed were categorized into three groups, strong inhibition: wild-type, intermediate inhibition: K17R/

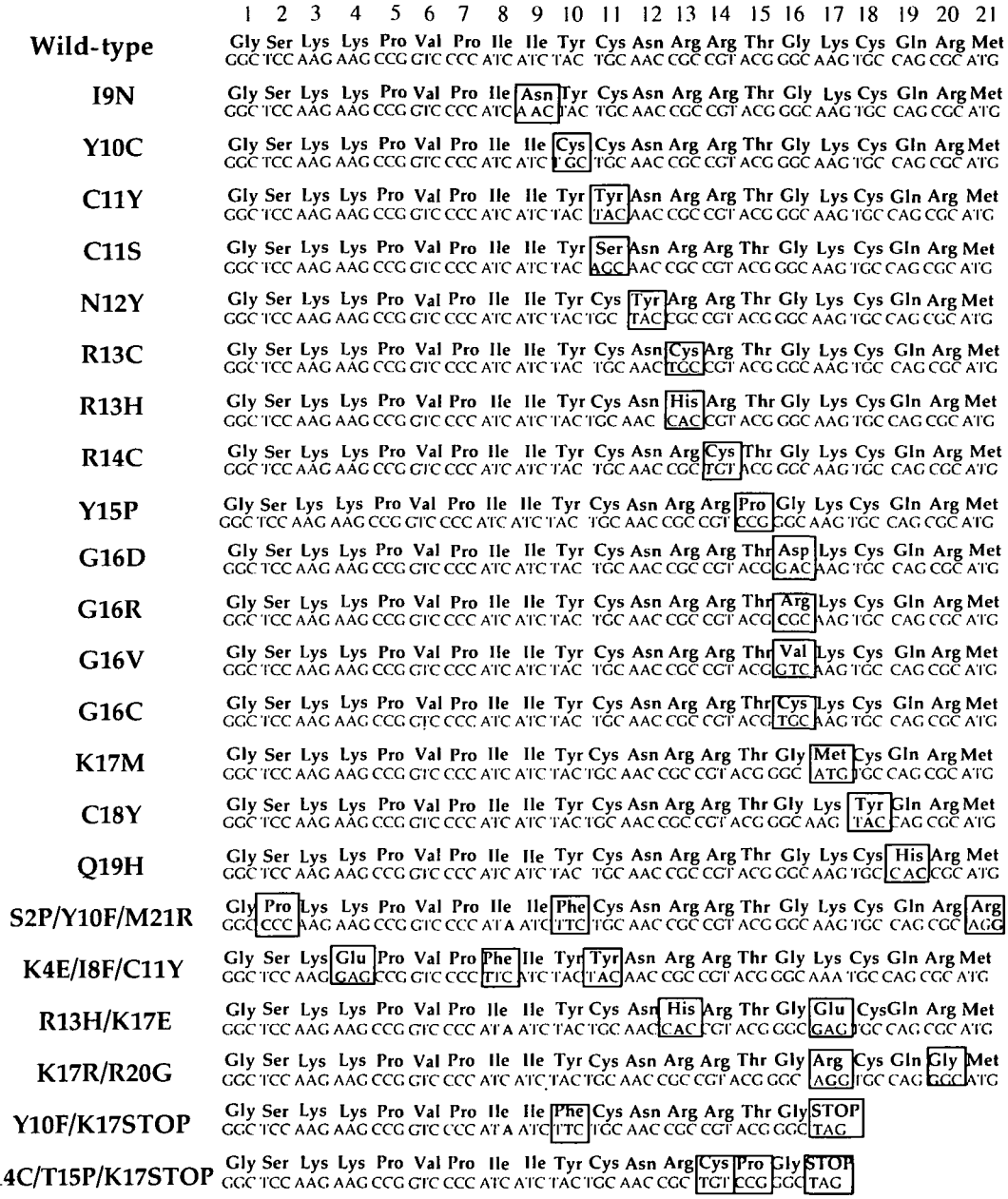


Fig. 2. **List of thanatin variants screened by the *in vivo* monitoring assay system.** Mutations in genes encoding thanatin variants were identified by nucleotide sequencing. Mutated nucleotides (mutated codons) are indicated by bold letters and consequently altered amino acid residues are boxed. Silent mutations were found at positions, Ile-8 (ATC→ATA) and Lys-17 (AAG→AAA).

R20G, and weak or nearly no inhibition: I9N, G16C, and G16D (slight activity detectable), and others, in good accordance with the results obtained in the primary plate assay. Thus, the spectrophotometric estimation of growth inhibition closely related to antimicrobial activity is considered to be useful as a secondary screening method for the precise confirmation of the primary screening. No significant variation in the amounts of SSI-tan fusions was observed among the variants tested, as judged by Western blot analysis using anti-SSI polyclonal antibody (data not shown).

Recombinant Overproduction of Thanatin—In order to evaluate whether the reduction in growth inhibition observed in many variants of thanatin was indeed due to mutation effects in the thanatin gene, we next attempted to isolate and characterize the variant peptides. Among the variants, three representative variants, K17R/R20G (intermediate activity), C11Y (remarkable activity loss), and M21R (remarkable activity loss), were chosen to study the structure–function relationship. C11Y and M21R are considered to be useful for understanding the role of disulfide bridge formation (C11Y) and the significance of C-terminal Met (M21R), which is known to be responsible for the antimicrobial activity. In establishing the production-purification system for the recombinant thanatin peptide, there are two conflicting problems; one is the need for overproduction for the detailed characterization of the thanatin peptide, and the other is the difficulty in obtaining a thanatin-producing *E. coli* transformant due to the high toxicity of the

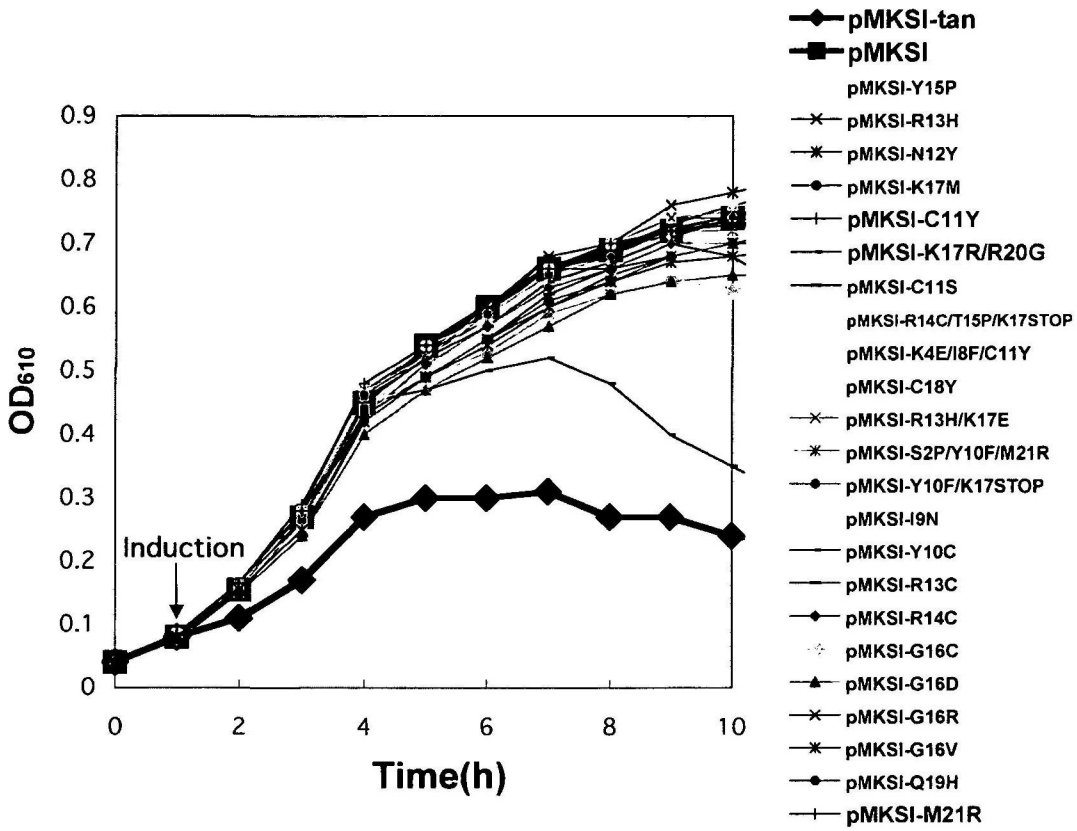


Fig. 3. **Growth curves of transformants producing thanatin variants fused with SSI.** Growth inhibition caused by induction (addition of 0.01 mM IPTG at OD_{610} = 0.1) was monitored spectrophotometrically for transformants harboring 23 plasmids with SSI-mutated thanatin fused genes, along with the SSI-wild-type thanatin fused gene, and the SSI gene alone. OD_{610}, optical density at 610 nm.

thanatin gene product, as experienced using the secretory expression plasmid vector pOS-tan, which led to lethal expression. Recombinant production as an inclusion body protein including thanatin was considered to a satisfactory solution to both problems. We constructed an overproduction system using pMTAKHV1 (*9*) for thanatin production by fusing the thanatin gene with the gene for a protein moiety (porcine adenylate kinase), thus causing the formation of inclusion bodies with large amounts of the fused protein, as shown in Fig. 1. In our previous study, the use of a Met-linker that would be cleaved with cyanogen bromide was very effective for the release of small-sized peptides such as apidaecin from the fused protein with SSI (*2*, *3*, *18–20*). However, this chemical reaction disrupts the anti-*E. coli* activity of thanatin by converting the essential C-terminal Met in thanatin to homoserine or homoserine lactone (*4*). As an alternative cleavage design, Asp and Pro codons were added as linker sites in front of the thanatin gene, making a cleavage site for 70% formic acid in the fused protein to be produced (Fig. 1). This cleavage strategy consists of sequential chemical and enzymatic digestions by 70% formic acid (*10*) and proline iminopeptidase (*11*), respectively. The electrophoretic pattern in Fig. 4 shows the time course for the production and accumulation of the wild-type thanatin-fused protein, corresponding to 36 kDa, in the insoluble fraction of transformant cells of a cytoplasmic Lon protease-defective strain, *E. coli* 21(DE3), harboring pAHS-tan. This indicates the successful high-level production of the thanatin-fused protein as stable inclusion bodies.

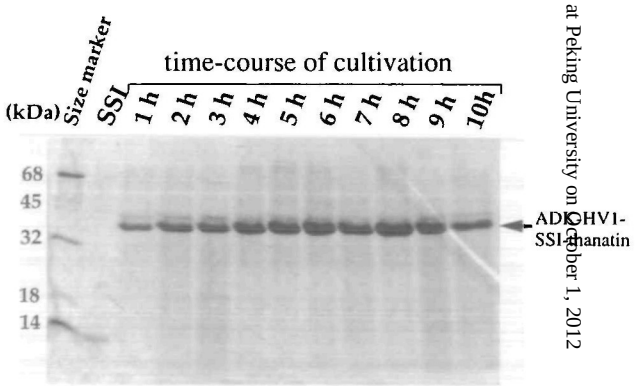


Fig. 4. **SDS-PAGE analysis of recombinant thanatin-fused protein.** An *E. coli* BL21(DE3) transformant harboring pAHS-tan was cultivated in 100 ml of M9 medium containing 0.2% glucose, 0.2% casamino acid, 1 mM thiamine-HCl, 1 mM $MgSO_4$–$7H_2O$, 0.1 mM $CaCl_2$, and ampicillin (100 μg/ml) at 37°C and induced with IAA at OD_{610} = 0.5, as described in "MATERIALS AND METHODS." After induction, 1 ml culture of *E. coli* BL21(DE3) transformant harboring pAHS-tan was harvested every 1 h (from 1 to 10 h). After measuring the cell number of the transformant, the harvested cells were sonicated to isolate the insoluble fraction including inclusion bodies containing mainly thanatin-fused protein, as described in "MATERIALS AND METHODS." Fractionated protein samples were precipitated by 8% TCA. Precipitates were resuspended in an adequate volume of TE buffer (10 mM Tris-HCl, pH 8.0) by adjusting to the same cell number (9.0×10^7 cells per 1 μl of TE buffer). Sample solutions (5 μl) were subjected to SDS-PAGE with SSI protein (2 μg).

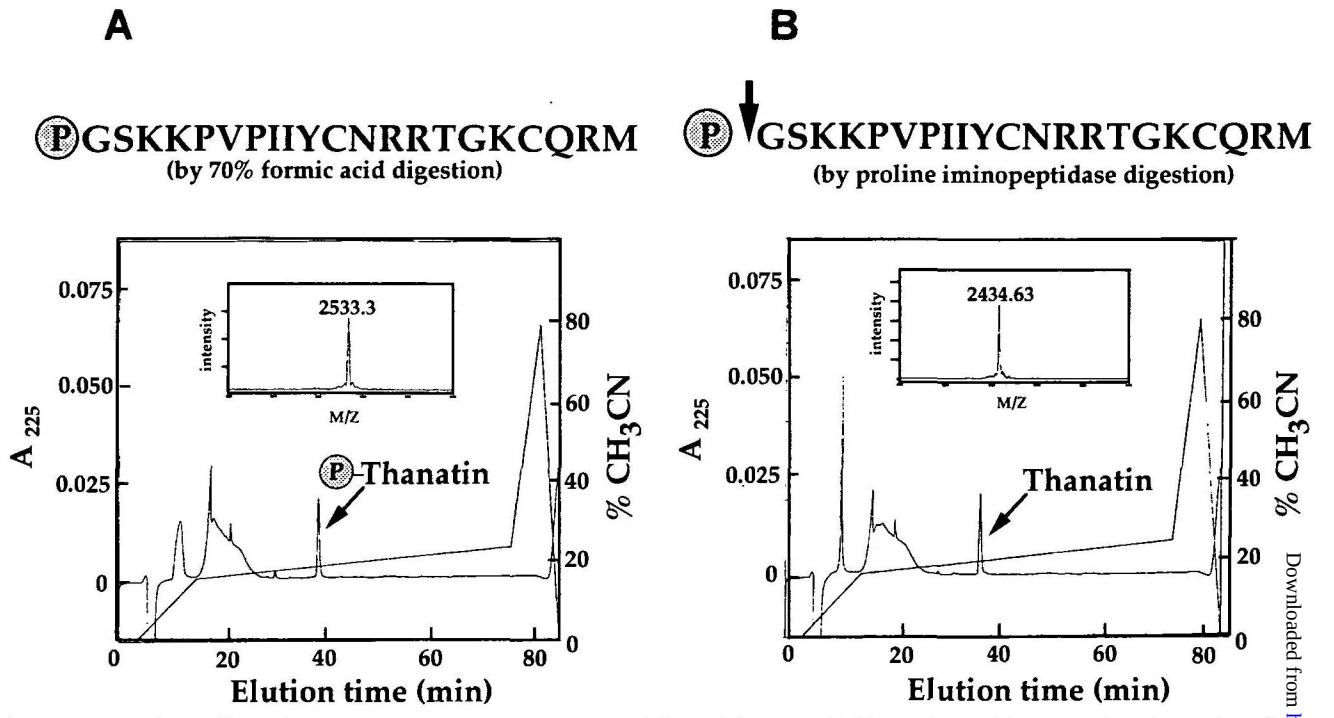


Fig. 5. **RP-HPLC profile and FAB mass spectrum.** (A) Pro-joined form of thanatin. (B) Native form of thanatin. Experimental conditions are described in detail in "MATERIALS AND METHODS."

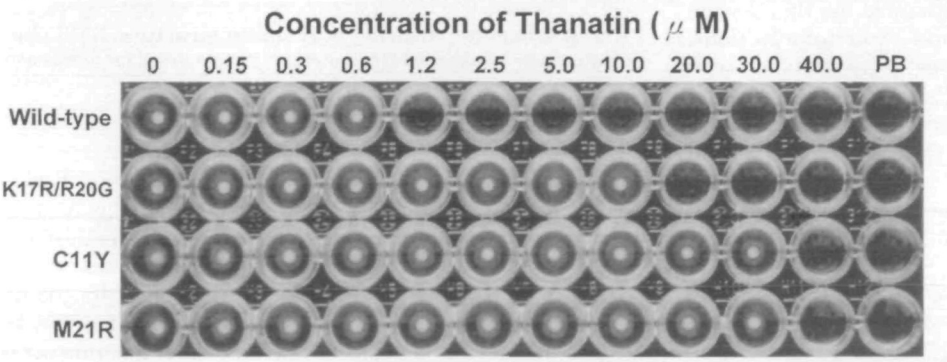


Fig. 6. **Antimicrobial activity assay.** Growth inhibition of *E. coli* JM109 was assayed by adding 20 μl of serial dilutes (concentrations indicated in the figure) of peptide samples into 80 μl of PB containing 300 cells of log-phase bacterial cells. From this liquid *in vitro* assay, the minimum inhibitory concentration (MIC) was determined for wild-type thanatin and its variants, K17R/R20G, C11Y, and M21R.

Purification and Characterization of Wild-Type and Variant Thanatins—The insoluble fraction (approximately 30 mg) including the thanatin-fused protein was concentrated by precipitation with 10% TCA and subjected to the reductive solution system, 8 M urea containing β-mercaptoethanol, and to 70% formic acid, in this order, with the aim of solubilizing the insoluble protein material and chemically releasing the Pro-joined thanatin portion from the fused protein. Formic acid catalyzed specifically at the site between the Asp and Pro preceeding the thanatin molecule, with total cleavage achieved in 12 h, as judged by SDS-PAGE using a tricine buffer system.

Figure 5A shows the RP-HPLC profile of the active fraction that exhibited strong antimicrobial activity against *E. coli* JM109. The bioactive peptide sample was subjected first to mass spectrometry and then to Edman degradation. FAB-MS yielded a mass of 2,533.3 Da (Fig. 5A). The first 17 amino acid residues of the isolated peptide were determined by Edman degradation, confirming the connection of a Pro residue to the thanatin peptide and the identical sequence (Gly1 to Gly16) to that predicted from the designed thanatin gene. The Pro-joined form was also used as one of the thanatin variants for the following *in vitro* bioassay. Removal of the Pro residue to gain the native form of thanatin was achieved by adding proline iminopeptidase to the Pro-joined thanatin peptide sample. The native thanatin had a slightly shorter retention time than the Pro-joined form due to the removal of hydrophobic Pro residue (Fig. 5B). The complete removal of the Pro residue was confirmed by protein sequence analysis. The molecular mass of the peptide (2,434.63 Da) was in good agreement with the calculated mass (2,436.2 Da), assuming that the two cysteine residues are engaged in an intramolecular disulfide bridge. The final yield of the thanatin peptide was calculated to be approximately 700 μg from 1 liter of culture. Throughout the purification process, a constant supply of argon gas into the reaction sample was very effective for avoiding undesired oxidation of the C-terminal Met residue

that has been reported to be important in antimicrobial action (*4*). In addition, three representative variants, C11Y, M21R (genetically generated) and K17R/R20G, were purified by the same method employed for the wild-type thanatin. All of the amino acid substitutions in these variants were confirmed by amino acid sequence and composition analysis.

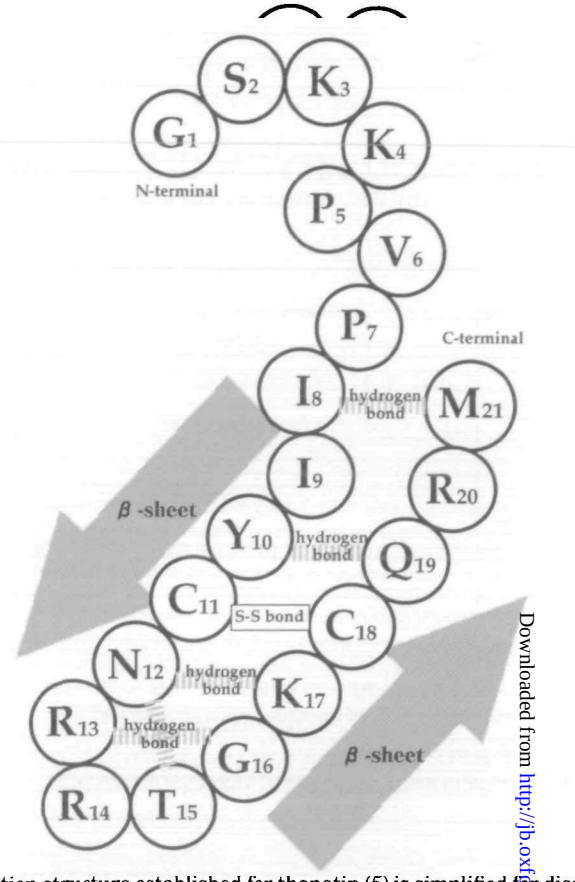
The purified recombinant wild-type and variant thanatins were examined for their antimicrobial activity against *E. coli* JM109, which was also used as the host cell for the *in vivo* assay. The results are shown in Fig. 6, in which bacterial growth can be observed at various concentrations (0–40 μM) of recombinant thanatin. The minimum inhibitory concentration (MIC) of wild-type thanatin lies between 0.6 and 1.2 μM, almost equal to the values obtained for other strains of *E. coli* used by Fehlbaum *et al.* (*4*). From these MIC values, the three representative thanatin variants were found to have reduced antimicrobial activities, K17R/R20G (20 μM), C11Y (40 μM), and M21R (40 μM). Therefore, it was concluded that the antimicrobial activity of each variant can be estimated conveniently on the basis of the growth curves.

Structure–Function Relationship of Thanatin—From the *in vivo* and *in vitro* assays (Figs. 3 and 6), the following discussion on structure–function relationship of thanatin can be presented, including the results of activity test experiments on several truncated isoforms of thanatin (*4*) and the NMR solution structure of thanatin (*5*). A schematic drawing of the thanatin structure (Fig. 7) based on the established tertiary architecture (*5*) is available for this discussion. Simply, thanatin comprises three structural portions, (i) a stretch of seven N-terminal mostly hydrophobic residues from G1 to P7, (ii) a cationic C-terminal loop engaged by a disufide bridge between C11 and C18, and (iii) C-terminal three-residue extension from Q19 to M21.

(a) Overall features: As for single mutations, it is notable that many anti–*E. coli*-activity-lowering mutations concentrate in the C-terminal half (Ile9 to Met21) of the peptide. In two variants, S2P/Y10F/M21R and K4E/I8F/C11Y, the reduced activities may be attributed at least to mutations at functionally important positions (*4*, *5*), M21R and C11Y, respectively, as discussed below. These results coincide well with the fact that truncations in the N-terminal region of the peptide are rather benign compared to those in the C-terminal region in keeping anti–*E. coli* activity with activity remaining in the case of deletions of three residues, Lys3 (100%), Pro5 (50%), and Pro7 (15%) (*4*). Such functional flexibility in the N-terminal region is also supported by our result that the Pro-joined form exhibits full activity (data not shown) and the NMR-based structural implication of a large structural variability of the N-terminus.

(b) M21R mutation: This mutation led to a remarkable reduction in activity, but slight activity against *E. coli* remained. The MIC value (40 μM) for the M21R variant was comparable to that (20–40 μM) for a variant lacking the last Met residue synthesized by Fehlbaum *et al.* (*4*). Cyanogen bromide treatment in 70% formic acid converted the C-terminal Met to homoserine or homoserine lactone, leading to drastically reduced activity (data not shown). It can be concluded unambiguously that the C-terminal Met is crucial for the antimicrobial action of thanatin against *E. coli*.

(c) C11Y mutation: Unexpectedly, this mutation did not result in a complete loss of activity. The disufide bridge (Cys11 and Cys18) was postulated to stabilize a well-defined β-sheet structure formed as the core of thanatin (*5*). However, the number of possible conformations is restricted to produce a slightly twisted formation by the disulfide bridging across anti-parallel β-strands. It might be argued whether a disufide bridge can be truly indispensable for the antimicrobial action, especially against *E. coli*, although its structural importance in maintaining the β-sheet structure is recognizable (*5*). Unfortunately, no amino acid replacement was performed within the C-terminal loop in previous truncation experiments (*4*). In our random mutagenesis experiment, all of the 14 single mutations within this structure resulted in drastically reduced activity. This strongly suggests the importance of a hydrogen bonding network holding the C-terminal loop together with a disufide bridge. Next, we attempted a molecular dynamics simulation under nearly natural conditions to explore the structural significance of the disufide bridge in the antibacterial action. As a result of MD for 1 ns, the β-sheet was relatively stable (Fig. 8A). The structures of the hydrophobic core and the side-chain of Arg20 were also stably maintained (*5*) (Fig. 8B). A large difference between the simulated and solution structure was seen in the N-terminal segment. It is not surprising to consider the fact that a large structural variability (18 models present) has been observed in the N-terminal region of the NMR structure (PDB; 8tfv) of thanatin (*5*). Because the β-sheet was de-



: solution structure established for thanatin (*5*) is simplified for discussion of the structure–function relationship.

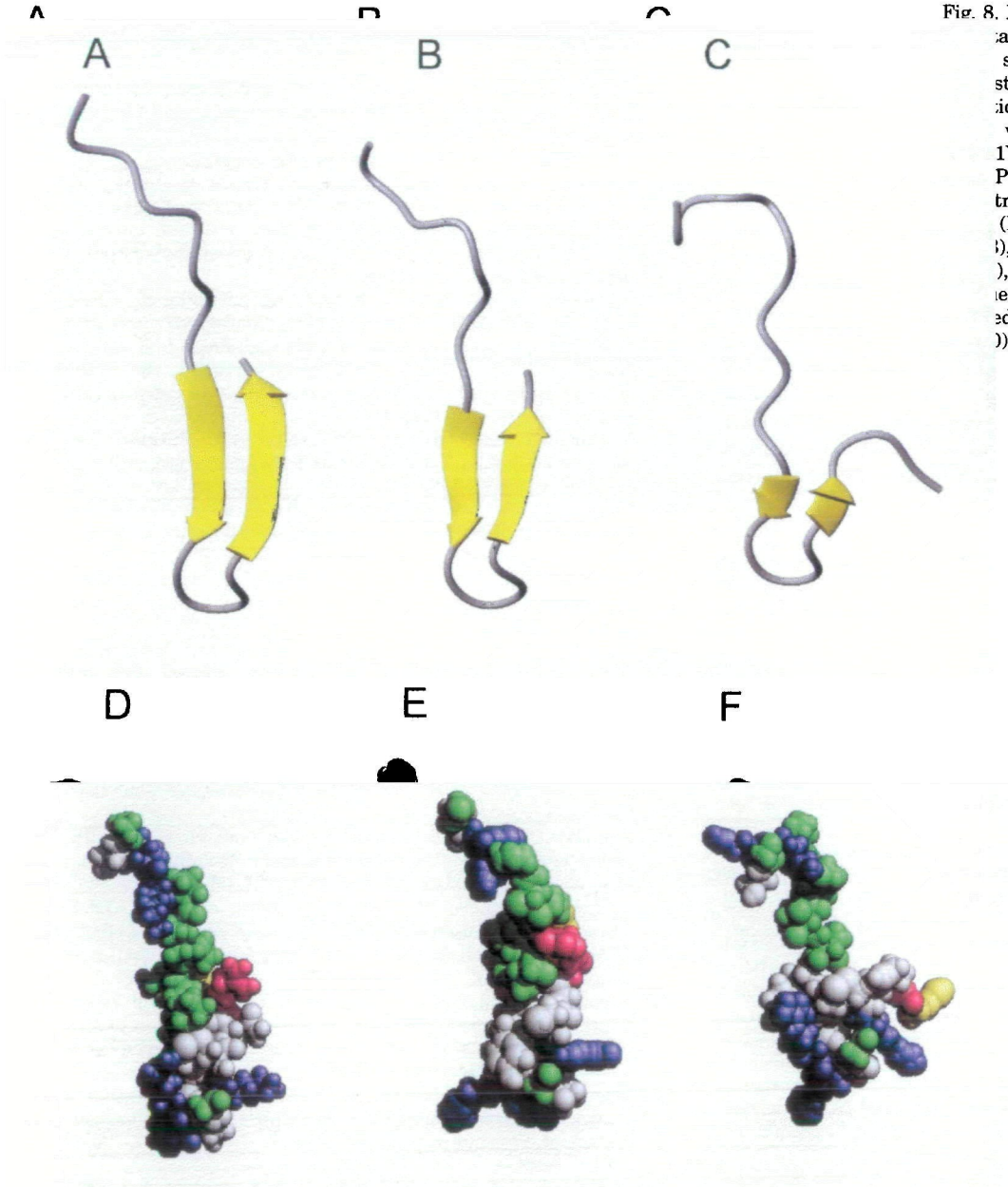


Fig. 8. **Molecular dynamics.** Rep- ation of simulated structures. structures are averaged over st 0.5 ns period. Ribbon repre- ion of (A); solution structure wild-type thanatin, (B); that 1Y, (C); that of K17R/R20G, PK representation of (D); solu- tructure of the wild-type tha- (E); C11Y, (F); K17R/R20G. In), and (C), β-sheet is in yellow.), (E), and (F), hydrophobic es are in green, positively ed residues are in blue, Arg20)) is in red and Met21 is in yel-

formed and formed repeatedly during 1 ns (data not shown), the disulfide bridge may contribute greatly to the formation and stabilization of the β-sheet in the peptide folding process. However, the S-S bond was found not to be indispensable for the functional expression since C11Y retains slight activity. In MD simulation, partial formation of the β-sheet structure occurred in this variant.

(d) K17R/R20G mutation: This variant with double mutations was selected because of its activity retention. Comparing both mutations, K17R and R20G, R20G would have a greater influence than K17R on thanatin activity, judging from the findings that the deletion of the C-terminal three residues including Arg20 leads to inactivation (*4*) and also from the importance of Arg20 in the tertiary structure of thanatin (*5*). In contrast, Arg17 can be speculated to be an alternate for Lys at this position, where a hydrogen bonding network is formed within the C-terminal loop. Each mutational contribution should be clarified by generating single mutants. Next, MD simulation was carried out on the K17R/R20G variant. MD for 1 ns showed that the β-sheet became shorter than in the solution structure (*5*) (Fig. 8, A and C). In the native state, Arg20 is involved in the hydrogen bonding network forming the β-sheet in the solution structure. However, the mutation of Arg20 to Gly gives rise to free rotation of the main-chain and consequently forces two C-terminal residues, Gly20 and Met21, to be separated from the hydrophobic core (Fig. 8, D and F). In such a situation, no significant mutational influence is detected in the N-terminal part of the hydrophobic core of the simulated structure of Met21. From the experimental evidence that the activity remains in this variant and M21R, it has been proven that the amino acid alteration of two residues at the C-terminal does not greatly affect the antimicrobial action. However, deletion of the three residues results in the loss of activity (*4*).

In conclusion, the structure–function relationship analy-

sis of thanatin based on the functional mapping gives data in good agreement with those obtained from deletion analysis with synthetic peptides performed by Fehlbaum *et al.* (*4*). This strategy provides useful detailed insights in understanding, in particular, the roles of amino acid residues involved in the hydrogen bonding network formed within the C-terminal loop. For further study on the molecular physiology of the antimicrobial peptide, it is necessary to consider the structural conformation related to activity under physiological conditions, where the antimicrobial peptide thanatin interacts with an unknown target molecule. Many variants isolated here would be useful for investigating the mode of antimicrobial action of thanatin not only against *E. coli* but also against gram-positive bacteria or fungi. Furthermore, a similar functional mapping technique should be established as a universal methodology that is widely applicable to expression systems using other microorganisms of interest.

We are very indebted to Dr. A. Tsugita and his coworkers, Science University of Tokyo, for their useful discussion and excellent performance of the FAB-MS and amino acid sequencing experiments.

REFERENCES

1. Hoffmann, J.A. (1995) Innate immunity of insects. *Curr. Opin. Immunol.* **7**, 4–10
2. Taguchi, S., Nakagawa, K., Maeno, M., and Momose, H. (1994) In vivo monitoring system for structure-function relationship analysis of the antibacterial peptide apidaecin. *Appl. Environ. Microbiol.* **60**, 3566–3572
3. Taguchi, S., Ozaki, A., Nakagawa, K., and Momose, H. (1996) Functional mapping of amino acid residues responsible for the antibacterial action of apidaecin. *Appl. Environ. Microbiol.* **62**, 4652–4655
4. Fehlbaum, P., Bulet, P., Chernysh, S., Briand, J.-P., Roussel, J.-P., Letellier, L., Hetru, C., and Hoffmann, J.A. (1996) Structure-activity analysis of thanatin, a 21-residue inducible insect defense peptide with sequence homology to frog skin antimicrobial peptides. *Proc. Natl. Acad. Sci. USA* **93**, 1221–1225
5. Mandard, N., Sodano, P., Labee, H., Bonmatin, J.-M., Bulet, P., Hetru, C., Ptak, M., and Vovelle, F. (1998) Solution structure of thanatin, a potent bactericidal and fungicidal insect peptide, determined from proton two-dimentional nuclear magnetic resonance data. *Eur. J. Biochem.* **256**, 404–410
6. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
7. Taguchi, S., Kumagai, I., and Miura, K. (1990) Comparison of secretory expression in *Escherichia coli* and *Streptomyces* of *Streptomyces* subtilisin inhibitor (SSI) gene. *Biochim. Biophys. Acta* **1049**, 278–285
8. Taguchi, S., Yoshida, Y., Matsumoto, K., and Momose, H. (1993) Improved leader and putative terminator sequences for high-level production of *Streptomyces* subtilisin inhibitor in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **39**, 732–737
9. Hibino, T., Misawa, S., Wakiyama, M., Maeda, S., Yazaki, K., Kumagai, I., Ooi, T., and Miura, K. (1994) High-level expression of porcine muscle adenylate kinase in *Escherichia coli*: effects of the copy number of the gene and the translational initiation signals. *J. Biotech.* **32**, 139–148
10. Kawakami, T., Kamo, M., Takamoto, K., Miyazaki, K., Chow, L.-P., Ueno, Y., and Tsugita, A. (1997) Bond-specific chemical cleavages of peptides and proteins with perfluoric acid vapors: Novel peptide bond cleavages of glycyl-threonine, the amino side of serine residues and the carboxyl side of aspartic acid residues. *J. Biochem.* **121**, 68–76
11. Yoshimoto, T. and Tsuru, D. (1985) Proline iminopeptidase from *Bacillus coagulans*: purification and enzymatic properties. *J. Biochem.* **97**, 1477–1485
12. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685
13. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935
14. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., and Kollman, P.A. (1995) A second generation force field for the simulation of protein and nucleic acids. *J. Am. Chem. Soc.* **117**, 5179–5197
15. Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatam, T.E., Cheng, A.L. *et al.* (1997) *AMBER5.0*, University of California, San Francisco
16. Berendsen, H.J.C., Postma, J.M.P., van Gunsteren, W.F., DiNola, A., and Haak, J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Comp. Phys.* **81**, 3684–3690
17. Ryckaert, J.-P., Ciccotti, G., and Berendsen, H.J.C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.* **23**, 327–341
18. Taguchi, S., Kumagai, I., Nakayama, J., Suzuki, A., and Miura, K. (1989) Efficient extracellular expression of a foreign protein in *Streptomyces* using secretory protease inhibitor (SSI) gene fusions. *Bio/technology*, **7**, 1063–1066
19. Maeno, M., Taguchi, S., and Momose, H. (1993) Production of antibacterial peptide 'apidaecin' using the secretory expression system of *Streptomyces*. *Biosci. Biotech. Biochem.* **57**, 1206–1207
20. Taguchi, S., Misawa, S., Yoshida, Y., and Momose, H. (1995) Microbial secretion of biologically active human transforming growth factor α fused to the *Streptomyces* protease inhibitor. *Gene* **159**, 239–243