THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 285, NO. 39, pp. 30150–30158, September 24, 2010

Isolation, cDNA Cloning, and Structure-based Functional Characterization of Oryctin, a Hemolymph Protein from the Coconut Rhinoceros Beetle, *Oryctes rhinoceros*, as a Novel Serine Protease Inhibitor*[S]

Received for publication, March 18, 2010, and in revised form, June 21, 2010 Published, JBC Papers in Press, July 14, 2010, DOI 10.1074/jbc.M110.124735

Shoichiro Horita[‡], Jun Ishibashi[§], Koji Nagata[‡], Takuya Miyakawa[‡], Minoru Yamakawa[§¶], and Masaru Tanokura[‡1]

From the [‡]Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan, the [§]National Institute of Agrobiological Sciences, Tsukuba, Ibaraki 305-8634, Japan, and the [¶]Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8572, Japan

We isolated oryctin, a 66-residue peptide, from the hemolymph of the coconut rhinoceros beetle *Oryctes rhinoceros* and cloned its cDNA. Oryctin is dissimilar to any other known peptides in amino acid sequence, and its function has been unknown. To reveal that function, we determined the solution structure of recombinant ^{13}C,^{15}N-labeled oryctin by heteronuclear NMR spectroscopy. Oryctin exhibits a fold similar to that of Kazal-type serine protease inhibitors but has a unique additional C-terminal α-helix. We performed protease inhibition assays of oryctin against several bacterial and eukaryotic proteases. Oryctin does inhibit the following serine proteases: α-chymotrypsin, endopeptidase K, subtilisin Carlsberg, and leukocyte elastase, with K_i values of 3.9×10^{-10} M, 6.2×10^{-10} M, 1.4×10^{-9} M, and 1.2×10^{-8} M, respectively. Although the target molecule of oryctin in the beetle hemolymph remains obscure, our results showed that oryctin is a novel single domain Kazal-type inhibitor and could play a key role in protecting against bacterial infections.

Eggs and larvae of the coconut rhinoceros beetle *Oryctes rhinoceros* live in compost in warm areas such as Southeast Asia. This beetle has thus developed a self-defense system that includes antimicrobial peptides. Several antimicrobial peptides were isolated from the hemolymph of the beetle: defensin (1), rhinocerosin (2), and scarabaecin (3). Oryctin (GenBank™ accession no. BAA36402), a 66-residue peptide with three intramolecular disulfide bonds, was found as a hemolymph peptide of the beetle during a search for antibacterial peptides. The function of oryctin has been unclear, because no other known peptide or protein is sequentially similar to oryctin. We took a structural approach to reveal the function of oryctin. First, we determined the solution structure of recombinant ^{13}C,^{15}N-labeled oryctin by heteronuclear NMR spectroscopy. Next, we searched for proteins that are structurally similar to oryctin in the Protein Data Bank (PDB).[2] The search revealed that oryctin has a fold that is similar in part to that of the turkey ovomucoid third domain (OMTKY3), a serine protease inhibitor. We then performed protease inhibition assays of oryctin and found that oryctin inhibits eukaryotic chymotrypsin-like serine proteases such as α-chymotrypsin and leukocyte elastase and bacterial subtilisin-like serine proteases such as subtilisin Carlsberg and endopeptidase K. Therefore, oryctin is a novel single domain Kazal-type inhibitor despite its unique amino acid sequence. Kazal-type serine protease inhibitors usually consist of multiple Kazal domains, each of which has a characteristic disulfide linkage pattern, Cys^{I}-Cys^{V} (where Cys^{I}, for example, is the first cysteine residue from the N terminus), Cys^{II}-Cys^{IV}, and Cys^{III}-Cys^{VI}, as well as a secondary structure consisting of an α-helix and an antiparallel β-sheet. Here, we discuss the structure and function of oryctin by comparing the sequences, the patterns of the disulfide linkages, and the tertiary structures.

EXPERIMENTAL PROCEDURES

Purification of Oryctin

The hemolymph of the third instar larvae of *O. rhinoceros* collected on the islands of Okinawa and Ishigaki, Japan, was collected into an ice-cooled 50-ml centrifugation tube containing 1 mg of aprotinin 24 h after injection of heat-killed *Escherichia coli*. The hemolymph was centrifuged at 39,100 × *g* for 50 min at 4 °C. The supernatant was heated in boiling water for 10 min and then cooled on ice and centrifuged at 39,100 × *g* for 50 min at 4 °C. The supernatant was acidified with 0.1% (v/v) trifluoroacetic acid (TFA) and applied onto a Sep-Pak Vac tC18

* This work was supported by the Enhancement of Center of Excellence, Special Coordination Funds for Promoting Science and Technology, Science and Technology Agency, Japan; and by the National Project on Protein Structural and Functional Analyses, the Targeted Proteins Research Program, and grants-in-aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

The atomic coordinates and structure factors (code 2KSW) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Table S1, Figs. S1–S6, and additional references.

[1] To whom correspondence should be addressed: Dept. of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan. Tel.: 81-3-5841-5165; Fax: 81-3-5841-8023; E-mail: amtanok@mail.ecc.u-tokyo.ac.jp.

[2] The abbreviations used are: PDB, Protein Data Bank; HSQC, heteronuclear single-quantum coherence; MCA, 4-methylcoumaryl-7-amide; MOCAc/Dnp, (7-methoxycoumarin-4-yl)-acetyl/2,4-dinitrophenyl; RACE, rapid amplification of cDNA end; NOESY, nuclear Overhauser enhancement spectroscopy.

column (Waters, Milford, MA) equilibrated with 0.1% (v/v) TFA and eluted with 10, 20, 30, 40, 50, and 100% (v/v) acetonitrile containing 0.1% (v/v) TFA. The 40% (v/v) acetonitrile fraction was applied to a PepRPC column (GE Healthcare) equilibrated with 0.05% (v/v) heptafluorobutanoic acid. The adsorbent was eluted with a linear gradient of acetonitrile: 0–20% (v/v) in 5 min, followed by 20–40% (v/v) in 40 min at a flow rate of 0.5 ml/min. The fractions containing oryctin were applied to the same system but using 0.05% (v/v) TFA instead of 0.05% (v/v) heptafluorobutanoic acid.

Peptide Sequencing

Peptide sequencing was performed using LF3400 (Beckman Coulter, Fullerton, CA) and Procise cLC492 (Applied Biosystems, Foster City, CA) protein sequencers.

Matrix-assisted Laser Desorption Ionization Time-of-flight Mass Spectrometry (MALDI-TOF MS)

MALDI-TOF MS was measured on a Voyager Linear spectrometer (Applied Biosystems). About 1 pmol of purified oryctin was dissolved in 1 μl of 0.1% (v/v) TFA. The sample solution was then mixed with saturated sinapinic acid solution in 50% (v/v) acetonitrile containing 0.1% (v/v) TFA directly on the target.

cDNA Cloning

The cDNA encoding oryctin was cloned using the following three-step PCR amplification.

Step 1: Reverse transcriptase-PCR—The fat body was collected 10 h after the injection of heat-killed *E. coli*. The poly(A)$^+$ RNA was purified from the fat body using a Quick Prep mRNA purification kit (GE Healthcare). The first-strand cDNA was synthesized from 1.32 μg of the poly(A)$^+$ RNA using a First-Strand cDNA synthesis kit (GE Healthcare). Using the first-strand cDNA as a template, RT-PCR was performed using the following degenerate primers: the sense primer O1F (17-mer), whose sequence was deduced from the amino acid sequence of Val^1–Asp^6 (5′-GTNCCNGTNGGN(AT)(CG)NG-A-3′) and the antisense primer O1R (20-mer), whose sequence was deduced from the amino acid sequence of Asn^{23}–Val^{29} (5′-AC(AGT)ATNCCYTTYTCNGGRTT-3′). A 45-cycle step-down protocol was used: denaturation at 94 °C for 8 min (first cycle) or 30 s (second and following cycles), annealing at 60 °C (initial five cycles), 55 °C (second five cycles), 50 °C (third five cycles), 45 °C (fourth five cycles), or 40 °C (final 25 cycles) for 30 s and polymerization at 72 °C for 30 s. The resultant 86-bp fragment was subcloned into a TA cloning vector (Invitrogen). DNA sequencing was performed using an ABI Prism model 373A DNA sequencer with the dye-terminator protocol (Applied Biosystems).

Step 2: 3′-Rapid Amplification of cDNA End—3′-RACE was done using the first-strand cDNA as the template, the antisense NotI adaptor primer 5′-AACTGGAAGAATTCGCGGC-3′, and the sense primer O2F (22-mer), whose sequence was derived from the result of Step 1: 5′-CTGTGAACCCAAAC-TATGCACC-3′. A 35-cycle PCR was used: denaturation at 94 °C for 8 min (first cycle) or 30 s (second and following cycles), annealing at 50 °C for 30 s, and polymerization at 72 °C for 30 s. The amplified fragment was subcloned and sequenced as described above.

Step 3: 5′-Rapid Amplification of cDNA End—5′-RACE was done using the 5′-RACE kit (Invitrogen). The first-strand cDNA was synthesized using the antisense primer OGSP1 (16-mer), whose sequence was derived from the result of Step 2: TTATGGACGTGGTGCA. The terminal deoxynucleotidyl transferase reaction was performed on ice for 1 h. The nested PCR was performed using the antisense primer OGSP2 (24-mer) TTTTTGATTGCTTCTTCACACTCG and the abridged anchor primer. A 40-cycle PCR was used: denaturation at 94 °C for 6 min (first cycle) or 1 min (second and following cycles), annealing at 50 °C for 30 s, and polymerization at 72 °C for 1 min. The amplified fragment was subcloned and sequenced as described above.

Determination of Disulfide Linkages

About 1 μg of oryctin was dissolved in 10 μl of 20 mM sodium bicarbonate (pH 8.5) containing 0.1 μg of lysyl endopeptidase (Wako, Tokyo, Japan) and endoprotease Asp-N (Roche Applied Science) and incubated at 37 °C for 4 h. The reaction mixture was applied directly to MALDI-TOF MS. α-Cyano-4-hydroxycinnamic acid was used as the matrix.

Gene Expression Analysis

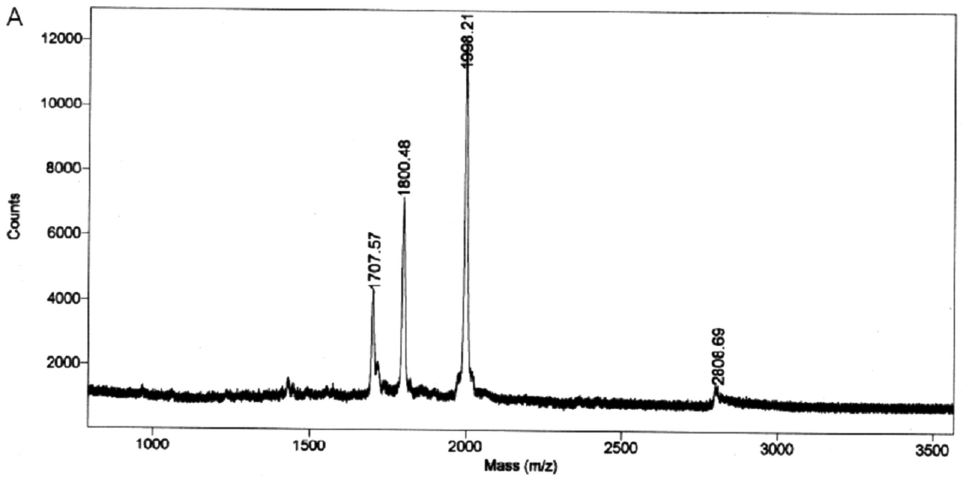
RT-PCR was carried out as described previously (1) using the following primers: 5′-AGTTGTGAATTTGTGCCGATA-CTG-3′ (forward) and 5′-CATTTGCGACAATTTATTTAA-AAGC-3′ (reverse). A 40-cycle PCR was used: denaturation at 94 °C for 6 min (first cycle) or 30 s (second and following cycles), annealing at 50 °C for 30 s, and polymerization at 72 °C for 30 s.

Expression and Purification of ^{15}N- and ^{13}C,^{15}N-labeled Oryctin

Oryctin with an N-terminal His_6 tag that was cleavable by tobacco etch virus (TEV) protease (Invitrogen) digestion was expressed in *E. coli* BL21-Star(DE3) (Invitrogen) grown in M9 minimal medium using a pET28a-based (Novagen, Madison, WI) expression plasmid by adding a final concentration of 1 mM isopropyl 1-thio-β-D-galactopyranoside at 37 °C for 3 h. Cells were harvested by centrifugation, resuspended in lysis buffer (50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 10 mM imidazole), and disrupted by sonication. After centrifugation, ^{15}N-labeled and ^{13}C,^{15}N-labeled oryctin with an N-terminal His_6 tag was purified with Ni Sepharose 6 Fast Flow (GE Healthcare). The His_6 tag was removed by tobacco etch virus (TEV) protease digestion. ^{15}N-labeled and ^{13}C,^{15}N-labeled oryctin were further purified by cation exchange followed by size exclusion chromatographies with Mono S HR 10/10 and Superdex 200 HR 10/30 (GE Healthcare), respectively.

NMR Spectroscopy

Purified ^{13}C,^{15}N-labeled oryctin was dissolved in 50 mM sodium phosphate buffer (pH 6.8), 100 mM NaCl, and 0.02% NaN_3 in 90% H_2O, 10% D_2O (v/v) to a final concentration of 2 mM. All NMR spectra were measured at 25 °C on a Unity Inova 500-MHz spectrometer (Varian, Palo Alto, CA) equipped with a triple-resonance probe. The following NMR data were acquired: ^{1}H-^{15}N HSQC, HNCO, HNCA, HN(CO)CA, CBCA-



B

VPVGSDC^{I}EPK	VGSDC^{I}EPK	LC^{II}TM	DLVPHC^{III}FLNPEK
C^{V}QNPNHEK	C^{V}QNPNHEK	GIVVVHGGC^{IV}ALSK	LGYTHEC^{VI}EEAIK
Calc. : 1998.22	Calc. : 1801.97	Calc. : 1705.12	Calc. : 2803.20
Meas. : 1998.21	Meas. : 1800.48	Meas. : 1707.57	Meas. : 2808.69

FIGURE 1. *A*, MALDI-TOF MS of natural oryctin after the digestion with lysyl endopeptidase and endoprotease Asp-N. *B*, structure and molecular mass (*m/z*) of the disulfide-linked fragments. N-terminally truncated oryctin (starting at Val^{3}) also existed in the sample. A site between Ser^{5} and Asp^{6} was not digested under this experimental condition.

(CO)NH, HNCACB, C(CO)NH, H(CCO)NH, HCCH-correlation spectroscopy (HCCH-COSY), HCCH-total correlation spectroscopy (HCCH-TOCSY), ^{15}N-edited NOESY-HSQC, and ^{13}C-edited NOESY-HSQC. The mixing times for the ^{15}N-edited and ^{13}C-edited NOESY-HSQCs were 75 and 150 ms. All of the acquired data were processed, visualized, and analyzed with the programs NMRPipe (4), NMRDraw (4), and Sparky (NMR Assignment and Integration Software, University of California, San Francisco), respectively.

Dihedral Angle Restraints

The programs CSI (5) and TALOS (6) were used to predict the regions of α-helix and β-strand based on $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, $^{1}H^{\alpha}$, and ^{15}N chemical shifts. The predicted angle ranges were used as the dihedral angle restraints.

Hydrogen Bond Restraints

Hydrogen bonds were detected by the following HgD exchange experiment. First, the ^{1}H-^{15}N HSQC spectrum of 0.2 mM ^{15}N-labeled oryctin dissolved in 50 mM sodium phosphate (pH 6.8), 100 mM NaCl, and 0.02% (w/v) NaN_3 in 90% H_2O, 10% D_2O (v/v) were acquired. Then, the solvent was changed to 50 mM sodium phosphate (pH 6.8), 100 mM NaCl, and 0.02% (w/v) NaN_3 in D_2O by ultrafiltration using a Vivaspin-20 (molecular weight cutoff 3000; Sartorius Stedim Biotech, Aubagne, France). ^{1}H-^{15}N HSQC spectra were acquired at 6, 12, 24, 36, and 48 h after the solvent exchange. The peak intensities in these ^{1}H-^{15}N HSQC spectra were analyzed. Slowly exchanging amide protons were detected, and their respective carbonyl acceptors were deduced from the NOE data, and the oryctin structures at intermediate stages of the structure calculations. Hydrogen bond distance restraints were set as $r_{HN\text{-}O} = 1.8–2.0$ Å, $r_{N\text{-}O} = 2.7–3.0$ Å.

Structure Calculation and Structural Similarity Search

Interproton distance restraints were derived from peak intensities in the ^{15}N-edited NOESY-HSQC and ^{13}C-edited NOESY-HSQC spectra of ^{13}C,^{15}N-labeled oryctin with a mixing time of 75 ms. The cross-peak intensities were translated into interproton distances based on the relationship, NOE $\propto$ (distance)$^{-6}$; the standard distance between H^{N}_{i} and H^{N}_{i+1} in the α-helix was 2.8 Å, and the standard distance between H^{α}_{i} and H^{α}_{j} in the antiparallel β-sheet was 2.3 Å. Structure calculations were performed using CYANA (7). Two hundred conformers were annealed in 10,000 steps of torsion angle dynamics calculations, of which 20 conformers with the lowest values in the target function were used to represent the solution structure of oryctin. The conformer with the lowest target function was used as the representative structure of oryctin. The tertiary structure was visualized with the programs MOLMOL (8) and PyMOL (DeLano, W. L.). The five conformers with the lowest target functions were submitted to the Dali server (9) to search for proteins structurally similar to oryctin.

Protease Inhibition Assays

α-Chymotrypsin, leukocyte elastase, subtilisin Carlsberg, endopeptidase K, trypsin, papain, and thermolysin were purchased from Sigma-Aldrich; V8 protease, pepsin, and lysyl endopeptidase were from Wako Pure Chemical Industries (Osaka, Japan); proline-specific endopeptidase was from Seikagaku Corp. (Tokyo, Japan); and endoproteinases Arg-C and Asp-N were from Roche Diagnostics. Peptidyl-MCA substrates (referred to as MCA substrates) and MOCAc/Dnp type fluorescence-quenching substrates (referred to as MOCAc/Dnp substrates) for the above-mentioned proteases (shown in Table 2) were purchased from Peptide Institute (Osaka, Japan). In the case of MCA substrates, the protease activity was measured using fluorescence intensity, where 380 and 460 nm were used as the excitation and detection wavelengths, respectively. In the case of MOCAc/Dnp substrates, the protease activity was measured using fluorescence intensity, where 328 and 393 nm were used as the excitation and detection wavelengths, respectively. The pH-dependent inhibitory activity was assayed as follows. A reaction mixture containing 3.3 nM protease, 10 nM oryctin, and 66, 100, or 130 μM MCA substrate or 6.6, 10, or 13 μM MOCAc/Dnp substrate in 10 mM Tris-HCl (pH 7.2–9.2) was incubated at 37 °C for 10 min, and then HCl was added to stop the reaction. Exceptionally, for a reaction involving pepsin,

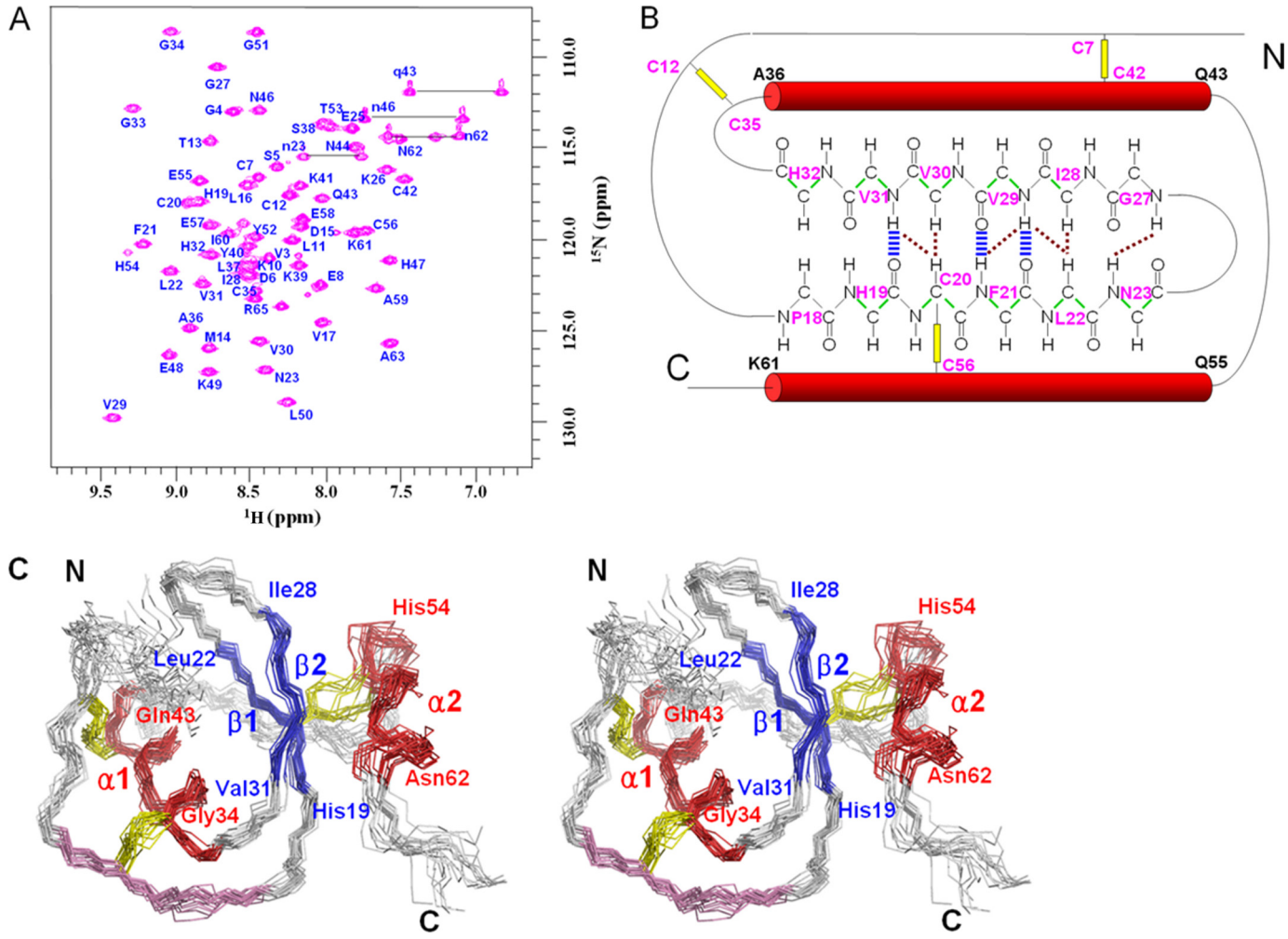


FIGURE 2. *A*, ^{1}H-^{15}N HSQC spectrum of oryctin. *B*, the secondary structure diagram of oryctin. The hydrogen bonds (*blue*) and interstrand NOEs (*brown*) in the antiparallel β-sheet are shown. The three disulfide bonds and the two α-helices are shown as *yellow lines* and *red tubes*. The dihedral angle restraints used are shown in *green*. *C*, stereoview of the superposition of 20 structures of oryctin. Two α-helices (α1, Gly^{34}–Gln^{43}; α2, His^{54}–Asn^{62}) and two β strands (β1, His^{19}–Leu^{22}; β2, Ile^{28}–Val^{31}) are shown in *red* and *blue*, respectively. The disulfide bonds are shown in *yellow*. The local sequence similarity between oryctin and OMTKY3 is shown in *purple*.

the pH of the reaction mixture was adjusted to 1.0–3.0 with HCl, and the reaction was stopped by adding 3 M Tris-HCl (pH 8.0). The temperature-dependent inhibitory activity was assayed as follows. A reaction mixture containing 3.3 nM protease, 10 nM oryctin, and 33, 66, 100, or 130 μM MCA substrate or 3.3, 6.6, 10, or 13 μM MOCAc/Dnp substrate in 10 mM Tris-HCl (pH 8.0) was incubated at 10–70 °C for 3 min, and then HCl was added to stop the reaction. The mode of protease inhibition and the inhibition constant (K_i) values were determined with a reaction mixture containing 3.3 nM protease, 83, 170, 250, 330, or 500 μM MCA substrate, and 0.83, 1.7, 2.5, 3.3 or 5.0 nM oryctin (for α-chymotrypsin, endopeptidase K, and subtilisin Carlsberg) or 8.3, 17, 25, 33, or 50 nM oryctin (for leukocyte elastase), respectively. The reaction mixtures were incubated at 37 °C for 5 min, and then HCl was added to stop the reaction. The acquired data were plotted on Lineweaver-Burk plots. The kinetics parameters (K_m and k_{cat}) were obtained with the reaction mixture containing 3.3 nM protease, 66, 100, or 130 μM MCA substrate or 6.6, 10, or 13 μM MOCAc/Dnp substrate in 10 mM Tris-HCl (pH 8.0). The reaction mixture was incubated at 37 °C for 3 min, and then HCl was added to stop the reaction. This reaction was performed in triplicate. Exceptionally, for a reaction involving pepsin, the pH of the reaction mixture was adjusted to 1.5 with HCl, and the reaction was stopped by adding 3 M Tris-HCl (pH 8.0).

Chemical Shift Perturbation

^{1}H-^{15}N HSQC spectra of [^{15}N]oryctin only, ^{15}N-labeled oryctin-α-chymotrypsin, and ^{15}N-labeled oryctin-leukocyte elastase were measured to investigate the protease recognition sites of oryctin. The final concentrations of ^{15}N-labeled oryctin and protease were 0.2 and 0.4 mM, respectively. The chemical shift change for each residue was determined quantitatively according to the function, $\delta\Delta = (\delta\Delta_{HN}{}^2 + 0.25\,\delta\Delta_{15N}{}^2)^{1/2}$ (10).

RESULTS

Peptide Isolation, cDNA Cloning, and Primary Structure Determination—Oryctin was found and isolated during the purification of antibacterial peptides from the hemolymph of the coconut rhinoceros beetle *O. rhinoceros*. About 1.0 μg of oryctin was obtained from 1.0 ml of the hemolymph by four steps of reverse-phase HPLC. The N-terminal amino acid sequence was determined using a protein sequencer, and the oryctin cDNA was cloned by three steps of PCR using fat body

mRNA from immunized larvae. The deduced amino acid sequence from the nucleotide sequence indicated that oryctin contains an 85-amino acid precursor, and a 66-amino acid mature peptide is assumed to be produced by cleavage of the signal peptide (supplemental Fig. S1). There is no sequence similarity between oryctin and any other known peptides or proteins. The MALDI-TOF MS data indicated that oryctin is a monomeric peptide with three intramolecular disulfide bonds and without any other modifications.

TABLE 1
Statistics for the 20 NMR structures of oryctin

Restraints	
Distance restraints derived from NOEs	
All	552
Intraresidual ($\|i-j\|=0$)	117
Sequential ($\|i-j\|=1$)	226
Medium range ($2 \le \|i-j\| \le 5$)	116
Long range ($\|i-j\|>6$)	103
Hydrogen bond restraints[a]	12
Disulfide bond restraints[b]	9
Dihedral angle restraints	
φ	38
ψ	41
Ramachandran plot (no. of residues/%)	
Most favored	815/76.9
Additionally allowed	228/21.5
Generously allowed	15/1.4
Disallowed	2/0.2
r.m.s.d. of atomic coordinates[c]	
Backbone atoms (Å)[d,e]	0.79 ± 0.17
Non-hydrogen atoms (Å)[e]	1.43 ± 0.21
Restraint violations	
r.m.s.d. distance restraint violations (Å)	0.018 ± 0.002
r.m.s.d. dihedral angle restraint violations	0.45 ± 0.09 °
Target function	
CYANA target function (Å^2)	1.3 ± 0.1

[a] Three hydrogen bonds each with four distance restraints.
[b] Three disulfide bonds each with three distance restraints.
[c] r.m.s.d., root mean square deviation.
[d] N, C$^\alpha$, and C′ of each residue.
[e] For well defined regions (residues 6–61).

The molecular masses (m/z) of detected ions after the digestion with lysyl endopeptidase and endoproteinase Asp-N were 1707.57, 1998.21, and 2808.69, respectively, using MALDI-TOF MS (Fig. 1). These ions were assigned to the pairs of Cys^{I}-Cys^{V}, Cys^{II}-Cys^{IV} and Cys^{III}-Cys^{VI}, respectively. An ion with an m/z of 1800.48 was assigned to fragments 3–10 and 42–49, which also indicated the disulfide bond Cys^{I}-Cys^{V}. Thus, we have concluded that three disulfide bonds are formed in oryctin: Cys^{7}–Cys^{42}, Cys^{12}–Cys^{35}, and Cys^{20}–Cys^{56} (Fig. 1).

Gene Expression Analysis—Oryctin was constitutively expressed in the fat bodies, hemocytes, midguts, and Malpighian tubules of the third instar larvae. Injection of *E. coli* into the larvae did not affect the expression of oryctin (supplemental Fig. S2).

NMR Structural Analysis—By using the conventional three-dimensional NMR data, 607 atoms of backbone and side chain ^{1}H,^{13}C,^{15}N were assigned. The ^{1}H-^{15}N HSQC spectrum of oryctin is shown in Fig. 2*A*. 100% (57 residues, except proline and the N-terminal residue) of the backbone ^{1}H,^{13}C,^{15}N atoms were assigned. The secondary structure regions predicted by both CSI and TALOS were restrained with the backbone dihedral angle (φ and ψ) restraints in the structure calculations (α, Ala^{36}–Gln^{43} and Gln^{55}–Lys^{61}; β, His^{19}–Leu^{22} and Val^{29}–Val^{31}). Three hydrogen bonds (Phe^{21} H^{N}–Val^{29} O, Val^{29} H^{N}–Phe^{21} O, and Val^{31}-H^{N}–His^{19} O) were assigned based on the low HgD exchange rates of amide protons and the oryctin structures at intermediate stages of structure calculations. A diagram showing secondary structures, hydrogen bonds, and interstrand NOEs is shown in Fig. 2*B*.

Tertiary Structure—The three-dimensional structure of oryctin was calculated based on 552 NOE-derived distance

TABLE 2
Kinetic parameters, specificity constants and inhibition constants of oryctin and aprotinin for the hydrolysis of MCA and MOCAc/Dnp substrates by eukaryotic and bacterial proteases

The parameters and constants were obtained at 37 °C and pH 8.0, except for pepsin (at 37 °C and pH 1.5). NI, no detectable inhibition.

Protease (origin)	Substrate	Without inhibitor			Inhibitor	K_i
		K_m	k_{cat}	k_{cat}/K_m		
		μM	s^{-1}	$s^{-1}\cdot m M^{-1}$		$n M$
α-Chymotrypsin (bovine)	Suc-AAPF-MCA[b]	110 ± 7.7	6.2 ± 0.54	58 ± 4.7	Oryctin/Aprotinin	0.39/1.3
Leukocyte elastase (human)	Suc-AAA-MCA[b]	82 ± 6.1	0.11 ± 0.007	1.4 ± 0.09	Oryctin/Aprotinin	12/1300
Subtilisin Carlsberg (*Butilus licheniformis*)	Suc-LLVY-MCA[b]	64 ± 5.8	2.5 ± 0.82	39 ± 4.5	Oryctin/Aprotinin	1.4/NI
Endopeptidase K (*Tritirachium album*)	Suc-AAPF-MCA[b]	84 ± 9.5	2.6 ± 0.35	31 ± 2.8	Oryctin/Aprotinin	0.62/NI
Trypsin (porcine)	Boc-QAR-MCA[b]	57 ± 0.29	5.7 ± 0.19	100 ± 5.1	Oryctin	NI
V8 protease (*Staphylococcus aureus*)	Z-LLE-MCA[b]	510 ± 52	0.43 ± 0.022	0.84 ± 0.044	Oryctin	NI
Proline-specific endopeptidase (*Flavobacterium meningosepticum*)	Suc-GP-MCA[b]	130 ± 6.9	0.041 ± 0.003	0.32 ± 0.015	Oryctin	NI
Lysyl endopeptidase (*Lysobacter enzymogenes*)	Boc-VLK-MCA[b]	15 ± 0.7	0.50 ± 0.049	32 ± 2.8	Oryctin	NI
Endoproteinase Arg-C (*Clostridium histolyticum*)	Boc-QAR-MCA[b]	190 ± 15	0.12 ± 0.009	0.63 ± 0.090	Oryctin	NI
Endoproteinase Asp-N (*Pseudomonas fragi*)	MOCAc-YVADAP-K(Dnp)-NH_2[b]	12 ± 0.1	53 ± 2.0	4400 ± 230	Oryctin	NI
Pepsin (porcine)	MOCAc-APAKFFRL-K(Dnp)-NH_2[b]	11 ± 0.4	370 ± 3.1	34000 ± 980	Oryctin	NI
Papain (*Papaya latex*)	Z-FR-MCA[b]	230 ± 16	0.22 ± 0.003	0.99 ± 0.028	Oryctin	NI
Thermolysin[a] (*Bacillus thermoproteolyticus* rokko)	MOCAc-KPLGL-A_2pr(Dnp)-AR-NH_2[b]	20 ± 2.2	380 ± 21	19000 ± 1200	Oryctin	NI

[a] The final concentration of 1 μM of zinc ion was added to activate proteolysis.
[b] Z, benzyloxycarbonyl; Suc, succinyl; Boc, *t*-butyloxycarbonyl; MOCAc, 7-methoxycoumarin-4-yl; MCA, 4-methylcoumaryl-7-amide; K(Dnp), N^{ϵ}-(2,4-dinitrophenyl)-L-lysine; and A_2pr(Dnp), N^{β}-(2,4-dinitrophenyl)-L-2,3-diaminopropionic acid.

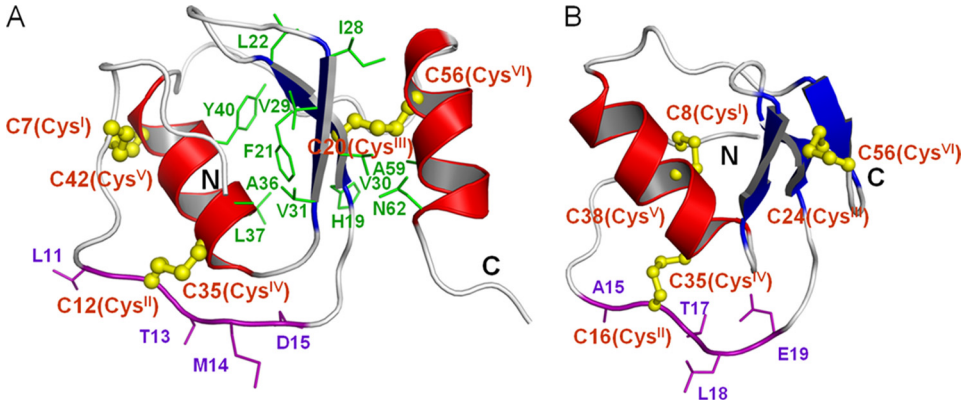


FIGURE 3. *A* and *B*, the ribbon representation of oryctin (*A*) and of OMTKY3 (*B*). The residues forming disulfide bonds and hydrophobic cores are shown in *yellow* and *green*, respectively. The local sequence similarity between oryctin and OMTKY3 is shown in *purple* (^{11}LCTMD^{15} (oryctin) *versus* ^{15}ACTLE^{19} (OMTKY3)).

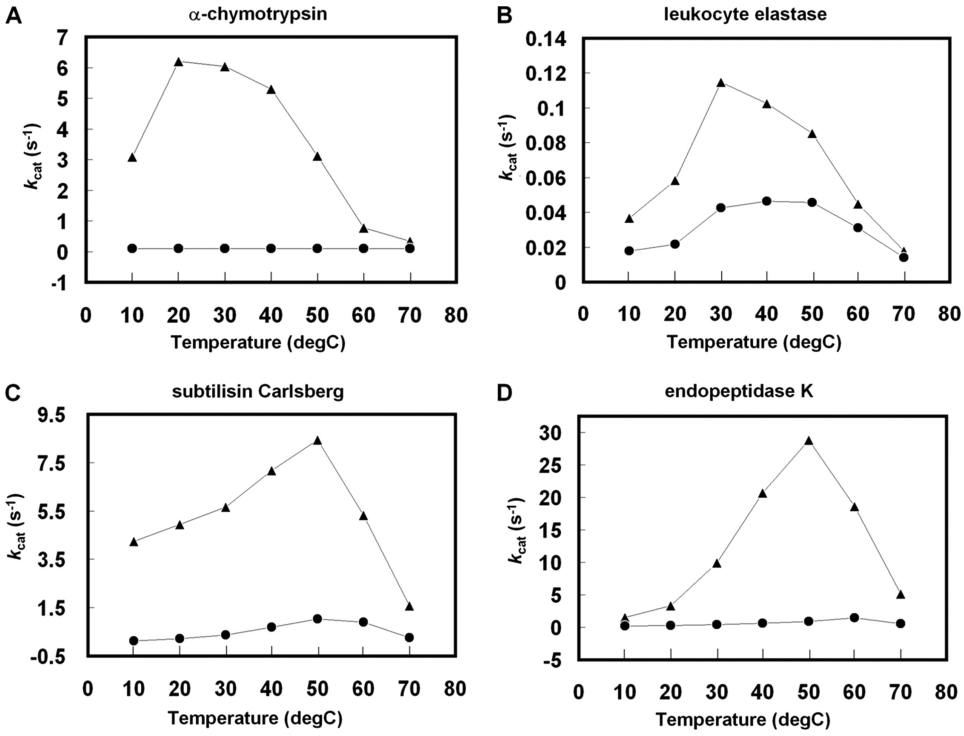


FIGURE 4. *A–D*, the temperature-dependent inhibition activity of oryctin toward α-chymotrypsin (*A*), leukocyte elastase (*B*), subtilisin Carlsberg (*C*), and endopeptidase K (*D*). The *triangles* and *circles* shown represent the turnover number (k_{cat}) in the absence and presence of oryctin (3-fold molar concentration compared with each protease), respectively.

restraints, three disulfide bond restraints (nine distance restraints), three hydrogen bond restraints (12 distance restraints), and 79 dihedral angle (38φ and 41ψ) restraints. A total of 200 structures were calculated, and the 20 structures with the lowest target functions were selected as an ensemble representing the solution structure of oryctin (Fig. 2*C*). The restraints used and the structural statistics for the final structure are summarized in Table 1. The oryctin molecule consists of a two-stranded antiparallel β-sheet (β1, His^{19}–Leu^{22}; β2, Ile^{28}–Val^{31}) and two α-helices (α1, Gly^{34}–Gln^{43}; α2, His^{54}–Asn^{62}). Three disulfide bonds were formed at Cys^{7}–Cys^{42}, Cys^{12}–Cys^{35}, and Cys^{20}–Cys^{56}. The pairwise root mean square deviations for backbone and all non-hydrogen atoms were 0.79 ± 0.17 Å and 1.43 ± 0.21 Å, respectively, among the 20 structures excluding five N- and five C-terminal residues. A representative structure with the lowest target function value is shown in Fig. 3*A*. The central β-sheet is rich in hydrophobic residues (His^{19}, Cys^{20}, Phe^{21}, Leu^{22}, Ile^{28}, Val^{29}, Val^{30}, and Val^{31}). The hydrophobic residues on the β-sheet and on the α-helices (α1, Ala^{36}, Leu^{37}, and Tyr^{40}; α2, Cys^{56}, Ala^{59}, and Asn^{62}) form a hydrophobic cluster and contribute to conformational stabilization. The Cys^{20}–Cys^{56} disulfide bond is included in the hydrophobic cluster.

Structural Similarity to the Third Domain of Turkey Ovomucoid—A structural similarity search using the Dali server (9) showed that oryctin is structurally similar to the third domain of OMTKY3 (turkey ovomucoid inhibitor) (Fig. 3*B*; root mean square deviation = 3.2 and 3.0 Å for 41 and 39 C$^{\alpha}$ pairs, respectively; PDB codes 1PPF and 1YU6, respectively), which is a Kazal-type serine protease inhibitor. Despite the lack of amino acid sequence similarity between them, their patterns of three disulfide bridges are the same: Cys^{I}-Cys^{V}, Cys^{II}-Cys^{IV} and Cys^{III}-Cys^{VI} (supplemental Fig. S3*A*). Oryctin, consisting of 66 residues, has three intramolecular disulfide bonds (Cys^{7}–Cys^{42}, Cys^{12}–Cys^{35}, and Cys^{20}–Cys^{56}), and OMTKY3, consisting of 51 residues, also has three intramolecular disulfide bonds (Cys^{8}–Cys^{38}, Cys^{16}–Cys^{35}, and Cys^{24}–Cys^{56}). In terms of secondary structure, oryctin has two α-helices and two β-strands, whereas OMTKY3 has one α-helix and three β-strands. Both oryctin and OMTKY3 possess a backbone fold similar to those of Kazal-type inhibitors: a central α-helix and an antiparallel β-sheet. The structural difference between oryctin and OMTKY3 is largest in the C-terminal regions, where oryctin has an amphipathic α-helix, whereas OMTKY3 has a short β-strand. The amino acid sequence of the reactive site loop (P4-P3-P2-P1-P1′) including Cys^{II} of OMTKY3 is Ala^{P4}-Cys$^{P3(II)}$-Thr^{P2}-Leu^{P1}-Glu$^{P1'}$, whereas that of oryctin is Leu^{P4}-Cys$^{P3(II)}$-Thr^{P2}-Met^{P1}-Asp$^{P1'}$ (shown in purple in Fig. 3, *A* and *B*), suggesting that oryctin also could function as a serine protease inhibitor.

Protease Inhibitory Activity—Supplemental Fig. S4 shows the pH-dependent inhibitory activity of oryctin against eukaryotic and bacterial proteases. Oryctin inhibited the eukaryotic serine proteases α-chymotrypsin and leukocyte elastase and the

bacterial serine proteases subtilisin Carlsberg and endopeptidase K over a pH range of 7.2–9.2, indicating it acts as a serine protease inhibitor in the beetle hemolymph. Fig. 4 shows the temperature-dependent inhibitory activity of oryctin against the above-mentioned four proteases, demonstrating that oryctin has inhibitory activity in a temperature range of 10–70 °C against the four proteases in 10 mM Tris-HCl (pH 8.0). Based on Lineweaver-Burk plots, oryctin inhibited α-chymotrypsin, endopeptidase K, subtilisin Carlsberg, and leukocyte elastase in a competitive manner with K_i values of 3.9×10^{-10} M, 6.2×10^{-10} M, 1.4×10^{-9} M, and 1.2×10^{-8} M, respectively (Fig. 5).

Putative Serine Protease Binding Region(s)—^{1}H-^{15}N HSQC spectra of ^{15}N-labeled oryctin only and ^{15}N-labeled oryctin-α-chymotrypsin showed similar distribution patterns of NMR signals. Chemical shift changes of oryctin by binding α-chymotrypsin and leukocyte elastase were evaluated quantitatively with the formula $\delta\Delta = (\delta\Delta_{HN}^{2} + 0.25\ \delta\Delta_{15N}^{2})^{1/2}$ (10) (Fig. 6, *A* and *B*). In the case of α-chymotrypsin binding (Fig. 6*A*), His[19], Cys[20], Gly[33], Gly[34], Leu[50], and His[54] exhibited large chemical shift changes (> 0.06 ppm); Thr[13], Val[31], His[32], Ala[36], Lys[49], Gly[51], Tyr[52], and Glu[58] exhibited moderate chemical shift changes (0.04–0.06 ppm); and Asp[15], Val[17], Val[29], Val[30], Ser[38], Glu[48], Glu[55], Glu[57], Ile[60], Asn[62], and Ala[63] exhibited small chemical shift changes (0.02–0.04 ppm). The residues with degenerated or invisible signals are not colored in this figure (Val[1], Val[3], Asp[6], Lys[10], Ile[28], Cys[35], and Lys[37]), nor are prolines (positions 2, 9, 18, 24, 45, 64, and 66). In the case of leukocyte elastase binding (Fig. 6*B*), His[19], Leu[50], His[54], and Glu[55] exhibited large chemical shift changes (> 0.06 ppm); Cys[20], Val[31], Gly[33], Gly[34], Tyr[52], and Cys[60] exhibited moderate chemical shift

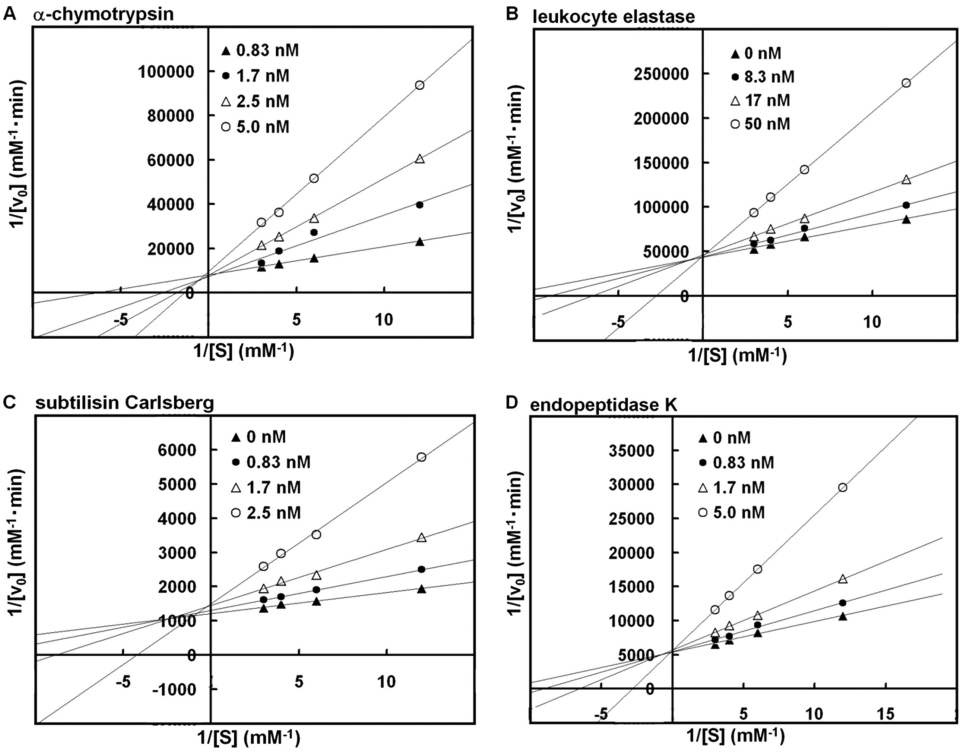


FIGURE 5. *A–D*, Lineweaver-Burk plot analysis of the inhibitory activity of oryctin against α-chymotrypsin (*A*), leukocyte elastase (*B*), subtilisin Carlsberg (*C*), and endopeptidase K (*D*). The inhibition constant (K_i) values against α-chymotrypsin, leukocyte elastase, subtilisin Carlsberg, and endopeptidase K were 3.9×10^{-10} M, 1.2×10^{-8} M, 1.4×10^{-9} M, and 6.2×10^{-10} M, respectively.

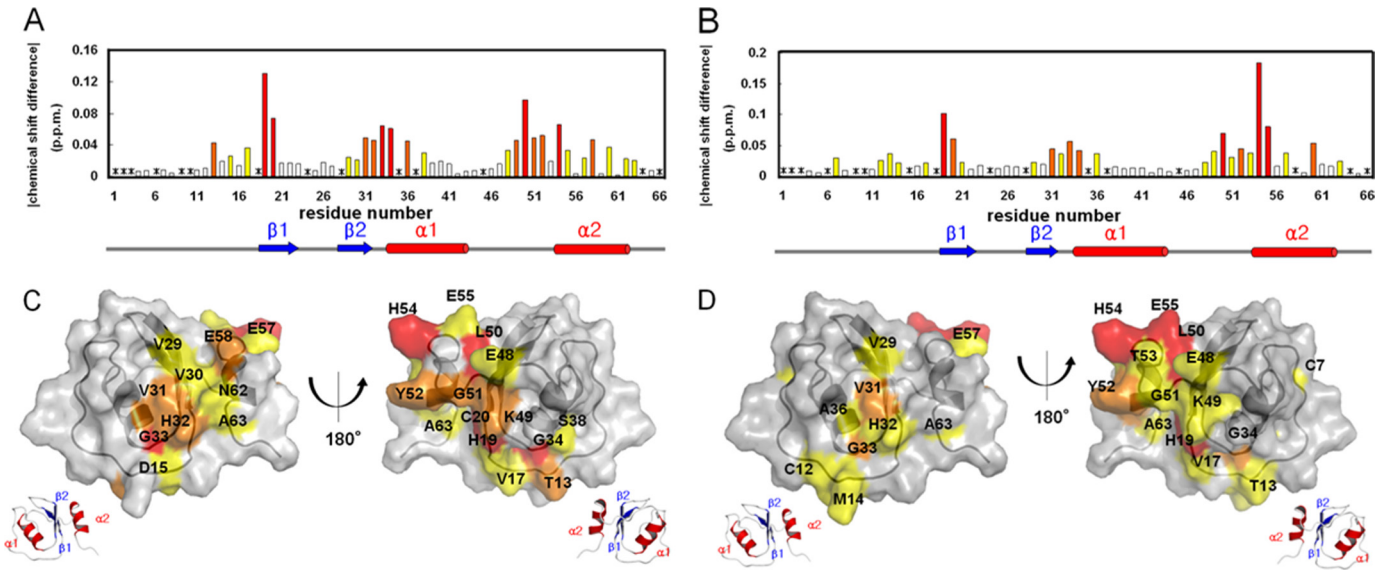


FIGURE 6. *A* and *B*, the chemical shift changes of oryctin between the free state and the bound state to α-chymotrypsin (*A*) and leukocyte elastase (*B*). *Asterisks* represent unassigned residues. *C* and *D*, mapping of the chemical shift-perturbed residues on the oryctin molecule by binding to α-chymotrypsin (*C*) and leukocyte elastase (*D*). The *bars* in (*A* and *B*) and the residues in (*C* and *D*) are colored according to the extent of chemical shift change defined by the formula, $\delta\Delta = (\delta\Delta_{HN}^{2} + 0.25\ \delta\Delta_{15N}^{2})^{1/2}$ (10). *White*, 0–0.02 ppm; *yellow*, 0.02–0.04 ppm; *orange*, 0.04–0.06 ppm; and *red*, >0.06 ppm.

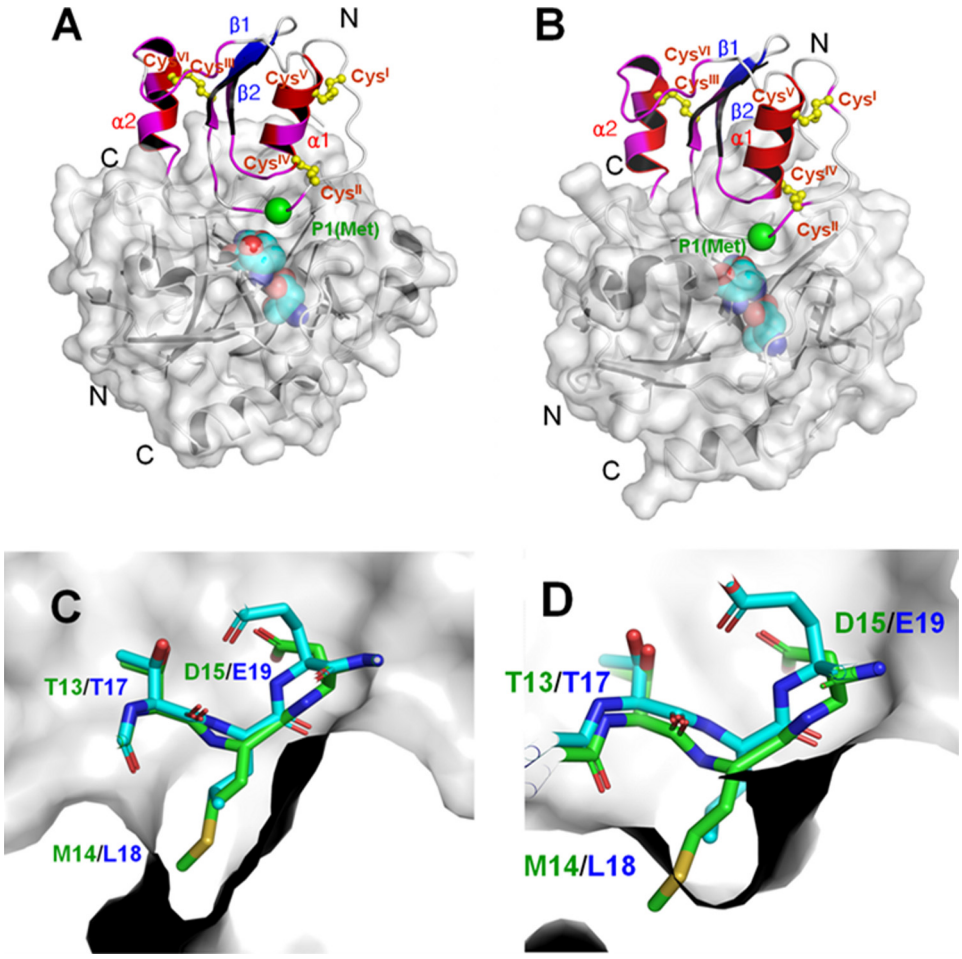


FIGURE 7. *A* and *B*, the binding model of oryctin to α-chymotrypsin (*A*) and leukocyte elastase (*B*), predicted based on the complex structures of the OMTKY3–α-chymotrypsin complex (PDB code 1CHO) and OMTKY3-leukocyte elastase complex (PDB code 1PPF) (supplemental Fig. S5). The chemical shift-perturbed residues (>0.02 ppm) are colored in *magenta*. The C^{α} atom of the reactive site residue (P1 residue) is shown as a *green sphere*. The disulfide bonds are shown in *yellow*. The active site triads of the α-chymotrypsin and leukocyte elastase are shown as *cyan spheres*. A unique additional α-helix at the C terminus (α2) of oryctin also would be involved in the recognition of α-chymotrypsin and leukocyte elastase and would contribute to the binding affinity. *C* and *D*, the crystal structure of the active site loop of OMTKY3 (*green*) and the model of the corresponding loop of oryctin (*blue*) in the S1 pockets of α-chymotrypsin (*C*) and leukocyte elastase (*D*). Met^{14}, the putative residue at the P1 position in oryctin, can fit into the S1 pocket of α-chymotrypsin, but it is too big to be accommodated by the S1 pocket of leukocyte elastase.

changes (0.04–0.06 ppm); and Cys^{7}, Cys^{12}, Thr^{13}, Met^{14}, Val^{17}, Phe^{21}, Val^{29}, His^{32}, Ala^{36}, Glu^{48}, Lys^{49}, Gly^{51}, Thr^{53}, Glu^{57}, and Ala^{63} exhibited small chemical shift changes (0.02–0.04 ppm). Degenerated or invisible signals are not colored in this figure (Val^{1}, Val^{3}, Asp^{6}, Lys^{10}, Asp^{15}, Ile^{28}, Cys^{35}, Lys^{37}, and Glu^{58}), nor are prolines (positions 2, 9, 18, 24, 45, 64, and 66). The amino acid residues whose chemical shifts were perturbed by the addition of α-chymotrypsin and/or leukocyte elastase are mostly common and located in three regions of the sequence (Fig. 6, *A* and *B*): (i) the putative reactive site near the N terminus (residues 13–20 and 12–21 in the cases of α-chymotrypsin and leukocyte elastase binding, respectively), (ii) the central part (residues 29–38 and 29–36), and (iii) the C-terminal part (residues 48–63 in both cases). When these residues are mapped on the oryctin structure, they form a circle around the molecule like a belt (Fig. 6, *C* and *D*). Because the similar surface areas of oryctin show chemical shift perturbations through the binding of oryctin to α-chymotrypsin or leukocyte elastase, oryctin should recognize these serine proteases in a very similar manner.

DISCUSSION

Proposed Inhibition Mode of Oryctin against α-Chymotrypsin and Leukocyte Elastase—The crystal structures of OMTKY3 bound to α-chymotrypsin and leukocyte elastase have already been solved (PDB codes 1CHO and 1PPF, respectively) (11, 12) (supplemental Fig. S5). OMTKY3 binds α-chymotrypsin and leukocyte elastase through its reactive site loop. The only sequentially similar segment between OMTKY3 and oryctin (^{15}ACTLE^{19} in OMTKY3 and ^{11}LCTMD^{15} in oryctin) is located in the active site loop of OMTKY3 and the corresponding loop in oryctin. Because oryctin inhibits chymotrypsin- and subtilisin-like family proteases in a competitive manner (Fig. 5), it should exhibit an inhibition mechanism similar to that of OMTKY3 by using the sequentially similar segment within the putative reactive site loop. Based on the crystal structure of OMTKY3 bound to α-chymotrypsin and leukocyte elastase, we have constructed a binding model of oryctin to α-chymotrypsin and leukocyte elastase (Fig. 7, *A* and *B*). In this model, not only the reactive site loop region but also the C-terminal α-helix uniquely found in oryctin is involved in the protease binding, which can cause a change in the orientation of the α-helix. This is consistent with the fact that the chemical shifts of most residues in the C-terminal α-helix are perturbed (> 0.02 ppm) by binding to the target proteases, although the chemical shift of the putative P1 residue, Met^{14}, is not perturbed.

α-Chymotrypsin and leukocyte elastase have large and small hydrophobic substrate-binding (S1) pockets and thus prefer large and small hydrophobic residues at the P1 position, respectively (Fig. 7, *C* and *D*). This explains why oryctin with Met at the P1 position inhibits α-chymotrypsin ($K_i = 3.9 \times 10^{-10}$ M) more tightly than leukocyte elastase ($K_i = 1.2 \times 10^{-8}$ M). The large S1 pocket of α-chymotrypsin will fit to the large hydrophobic side chain of Met^{14} at the P1 position, but the small S1 pocket of leukocyte elastase will be too small to accommodate the Met side chain. In contrast, OMTKY3 with Leu at the P1 position can bind both α-chymotrypsin and leukocyte elastase very tightly, with K_d values of 5.5×10^{-12} M and 2.4×10^{-11} M, respectively (11, 13), because the small side chain of Leu at the

P1 position can be accommodated in both the S1 pockets of α-chymotrypsin and leukocyte elastase.

The Number of Residues between Cys^{II} and Cys^{III} Provides Insight into the Functional Characterization of Kazal-type Inhibitors—Oryctin and the reported Kazal-type inhibitors that contain three disulfide bonds show no sequence similarity, indicating that oryctin and the other Kazal-type inhibitors have evolved convergently, whereas all of the Kazal-type inhibitors except oryctin have evolved divergently (supplemental Table S1 and supplemental Fig. S3*B*). In fact, the phylogenetic analysis of all the Kazal-type inhibitors in supplemental Table S1 except oryctin can be performed without error using the Phylogeny.fr server (supplemental Fig. S6) (14), but that with oryctin results in an error. Supplemental Table S1 compares the number of residues among six cysteine residues (Cys^{I}-Xn_a-Cys^{II}-Xn_b-Cys^{III}-Xn_c-Cys^{IV}-Xn_d-Cys^{V}-Xn_d-Cys^{VI}). The putative reactive site loop (Leu^{P4}-$Cys^{P3(II)}$-Thr^{P2}-Met^{P1}-$Asp^{P1'}$) of oryctin contains $Cys^{P3(II)}$. Met^{P1}, probably the most important for the inhibitory activity, is located between Cys^{II} and Cys^{III}. The number of residues between Cys^{II} and Cys^{III} (seven) is strictly conserved in all the members, including oryctin, with the exception of the second domain of rhodniin (15), which contains eight amino acid residues in that region. The number of residues between Cys^{II} and Cys^{III} yields insight into the functional characterization of the Kazal-type inhibitors. Interestingly, there are more residues between Cys^{I} and Cys^{II} in vertebrate inhibitors than in invertebrate ones (supplemental Table S1). The lymphoepithelial Kazal-type-related inhibitor (LEKTI) isolated from human blood ultrafiltrate, which consists of 15 Kazal domains, contains longer sequence stretches of 12–13 residues between Cys^{I} and Cys^{II} (16). This may imply that the N-terminal region of the Kazal-type inhibitors in higher animals played some additional role in the course of evolution.

Possible Biological Roles of Oryctin—Proteases and protease inhibitors are involved in the metabolism, immunity, and metamorphosis of insects (17, 18). In particular, hemolymph proteases and protease inhibitors are involved in immune responses in insects, *e.g.* antimicrobial peptide induction, prophenoloxidase activation, hemolymph coagulation, and protection against virulence-related microbial proteases (17). In the present study, we have shown that oryctin inhibits subtilisin-like serine proteases such as subtilisin Carlsberg and endopeptidase K. This suggests that oryctin is also involved in protection against microbial proteases. In the silkworm *Bombyx mori*, a protease inhibitor against fungal protease was isolated and is considered to play a role in innate immunity (19). Although further investigation is required to reveal the *in vivo* function(s) of oryctin, the results obtained in this study indicate that oryctin is a novel and unique member of Kazal-type serine protease inhibitors. Despite the lack of sequence similarity, oryctin exhibits a similar linkage pattern of three disulfide bonds and an inhibition mechanism similar to that of the traditional Kazal-type inhibitors.

Acknowledgments—We thank Jun Yang and Masanori Yamamoto for technical assistance.

REFERENCES

1. Ishibashi, J., Saido-Sakanaka, H., Yang, J., Sagisaka, A., and Yamakawa, M. (1999) *Eur. J. Biochem.* **266,** 616–623
2. Yang, J., Yamamoto, M., Ishibashi, J., Taniai, K., and Yamakawa, M. (1998) *Eur. J. Biochem.* **255,** 734–738
3. Tomie, T., Ishibashi, J., Furukawa, S., Kobayashi, S., Sawahata, R., Asaoka, A., Tagawa, M., and Yamakawa, M. (2003) *Biochem. Biophys. Res. Commun.* **307,** 261–266
4. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR* **6,** 277–293
5. Wishart, D. S., and Sykes, B. D. (1994) *J. Biomol. NMR* **4,** 171–180
6. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J. Biomol. NMR* **13,** 289–302
7. Jee, J., and Güntert, P. (2003) *J. Struct. Funct. Genomics* **4,** 179–189
8. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graph.* **14,** 51–55, 29–32
9. Holm, L., Kääriäinen, S., Rosenström, P., and Schenkel, A. (2008) *Bioinformatics* **24,** 2780–2781
10. Takeda, M., Terasawa, H., Sakakura, M., Yamaguchi, Y., Kajiwara, M., Kawashima, H., Miyasaka, M., and Shimada, I. (2003) *J. Biol. Chem.* **278,** 43550–43555
11. Fujinaga, M., Sielecki, A. R., Read, R. J., Ardelt, W., Laskowski, M., Jr., and James, M. N. (1987) *J. Mol. Biol.* **195,** 397–418
12. Bode, W., Wei, A. Z., Huber, R., Meyer, E., Travis, J., and Neumann, S. (1986) *EMBO J.* **5,** 2453–2458
13. Baker, B. M., and Murphy, K. P. (1997) *J. Mol. Biol.* **268,** 557–569
14. Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V., Lescot, M., Claverie, J. M., and Gascuel, O. (2008) *Nucleic Acids Res.* **36,** W465–469
15. Friedrich, T., Kröger, B., Bialojan, S., Lemaire, H. G., Höffken, H. W., Reuschenbach, P., Otte, M., and Dodt, J. (1993) *J. Biol. Chem.* **268,** 16216–16222
16. Lauber, T., Schulz, A., Schweimer, K., Adermann, K., and Marx, U. C. (2003) *J. Mol. Biol.* **328,** 205–219
17. Kanost, M. R. (1999) *Dev. Comp. Immunol.* **23,** 291–301
18. Natori, S., Shiraishi, H., Hori, S., and Kobayashi, A. (1999) *Dev. Comp. Immunol.* **23,** 317–328
19. Eguchi, M., Itoh, M., Nishino, K., Shibata, H., Tanaka, T., Kamei-Hayashi, K., and Hara, S. (1994) *J. Biochem.* **115,** 881–884