Distinct Carbohydrate Recognition Domains of an Invertebrate Defense Molecule Recognize Gram-negative and Gram-positive Bacteria*

Martin Bilej‡, Patrick De Baetselier§, Els Van Dijck§, Benoit Stijlemans§, Alain Colige¶, and Alain Beschin§‖

From the ‡Department of Immunology, Institute of Microbiology, Academy of Sciences of the Czech Republic, Prague 4, Czech Republic, the §Department of Immunology, Parasitology, and Ultrastructure, Flemish Interuniversity Institute for Biotechnology, Free University of Brussels, Sint-Genesius-Rode 1640, Belgium, and the ¶Laboratory of Connective Tissue Biology, University of Liège, Liège 4000, Belgium

Coelomic fluid of *Eisenia foetida* earthworms (*Oligochaeta, Annelida*) contains a 42-kDa defense molecule named CCF for coelomic cytolytic factor. By binding microbial antigens, namely the *O*-antigen of lipopolysaccharide (LPS), β-1,3-glucans, or *N,N'*-diacetylchitobiose present, respectively, on Gram-negative bacteria or yeast cell walls, CCF triggers the prophenoloxidase activating pathway. We report that CCF recognizes lysozyme-predigested Gram-positive bacteria or the peptidoglycan constituent muramyl dipeptide as well as muramic acid. To identify the pattern recognition domains of CCF, deletion mutants were tested for their ability to reconstitute the prophenoloxidase cascade in *E. foetida* coelomic fluid depleted of endogenous CCF in the presence of LPS, β-1,3-glucans, *N,N'*-diacetylchitobiose, and muramic acid. In addition, affinity chromatography of CCF peptides was performed on immobilized β-1,3-glucans or *N,N'*-diacetylchitobiose. We found that the broad specificity of CCF for pathogen-associated molecular patterns results from the presence of two distinct pattern recognition domains. One domain, which shows homology with the polysaccharide and glucanase motifs of β-1,3-glucanases and invertebrate defense molecules located in the central part of the CCF polypeptide chain, interacts with LPS and β-1,3-glucans. The C-terminal tryptophan-rich domain mediates interactions of CCF with *N,N'*-diacetylchitobiose and muramic acid. These data provide evidence for the presence of spatially distinct carbohydrate recognition domains within this invertebrate defense molecule.

Invertebrate innate defense strategies are based on pattern recognition receptors that do not discriminate between individual antigens but recognize surface determinants common to potential pathogens (1). Pattern recognition receptors interacting with different saccharide moieties of microbial lipopolysaccharides (LPS),[1] peptidoglycans, and β-1,3-glucans were described in numerous invertebrate and vertebrate species. We have characterized a 42-kDa lectin named coelomic cytolytic factor 1 (CCF) from the coelomic fluid of *Eisenia foetida* earthworms (*Oligochaeta, Annelida* (2, 3)). CCF displays significant amino acid sequence homology with bacterial and animal β-1,3-glucanases but does not exhibit such enzymatic activity (3–5). More interestingly, CCF shows homology with the α subunit of the β-1,3-glucan sensitive factor G from the horseshoe crab *Tachypleus tridentatus* (6) and with the Gram-negative bacteria-binding proteins of various insects (7–10). More recent reports concern the sequence homology of CCF with LPS and β-1,3-glucan recognition proteins from arthropods (11–13, for review see Ref. 14). All these invertebrate homologs have been suggested to play a role in invertebrate innate immunity acting as pattern recognition receptors. Accordingly, CCF binds with cell wall components of Gram-negative bacteria or yeast (the *O*-antigen of LPS, β-1,3-glucans, *N,N'*-diacetylchitobiose), triggering the activation of the prophenoloxidase (pro-PO) cascade (3, 15). This pathway results in the formation of cytotoxic and antimicrobial compounds and thus represents an important defense mechanism in a variety of invertebrates (for review see Refs. 16–18).

Pro-PO cascade is activated within 6 h in *E. foetida* coelomic fluid by Gram-negative but not Gram-positive bacteria (3). However, *E. foetida* coelomic fluid incubated for 6 h with Gram-negative or Gram-positive bacteria exhibits a similar antibacterial effect (19). Whether the failure of Gram-positive bacteria to induce the pro-PO cascade resulted from insufficient activation time is not clear.

The cell wall of virtually all bacteria consists of a rigid peptidoglycan layer that is either overlaid by an outer LPS layer in Gram-negative bacteria or remains exposed on the surface of Gram-positive bacteria. Peptidoglycan is a polymer of alternating *N*-acetylglucosamine and *N*-acetylmuramic acid units connected by short pentapeptides. The β-1,4-glycosidic bond of the *N*-acetylglucosamine-*N*-acetylmuramic acid pepti-

[1] The abbreviations used are: LPS, lipopolysaccharide; CCF, coelomic cytolytic factor 1; ECCF, *E. foetida* CCF; E1–E5, *E. foetida* CCF fragments 1–5; LCCF, *L. terrestris* CCF; L2 or L4, *L. terrestris* CCF fragments 2 or 4; pro-PO, prophenoloxidase; PO, phenoloxidase; MDP, muramyl dipeptide; CF, coelomic fluid proteins; kb, kilobase(s); RACE, rapid amplification of cDNA ends; GSP, generated specific primer; PCR, polymerase chain reaction; L-DOPA, 3-(3,4-dihydroxylphenyl)-L-alanine.

doglycan backbone can be hydrolyzed by lysozyme (muramidase; mucopeptide *N*-acetylmuramoylhydrolase; EC 3.2.1.17), a ubiquitous enzyme involved in innate immune reaction of numerous animal species, including annelids (20, 21).

In the present work, we examined whether Gram-positive bacteria trigger the pro-PO cascade in the earthworm coelomic fluid. We report that digestion of the Gram-positive bacteria or peptidoglycan by lysozyme-like activity makes it possible to induce the pro-PO cascade through the recognition of muramic acid by CCF. We also show that the broad specificity of CCF results from the presence of two distinct lectin-like domains within the molecule, a domain located in the central part of CCF implicated in interactions with LPS and β-1,3-glucans, and a C-terminal tryptophan-rich domain interacting with *N,N'*-diacetylchitobiose and muramic acid. Finally we observed that the CCF homolog in the *E. foetida*-related species *Lumbricus terrestris* displays distinct saccharide specificities and does not recognize *N,N'*-diacetylchitobiose and muramic acid. *E. foetida* CCF is, to our knowledge, the first invertebrate defense molecule acting as pattern recognition molecule for cell wall components of yeast and Gram-negative and Gram-positive bacteria.

EXPERIMENTAL PROCEDURES

Materials—Cell-free coelomic fluid was collected from adult specimens of *E. foetida foetida* (common brandling worm) and *L. terrestris* (common earthworm) in the absence of serine protease inhibitor as described previously (2). *Micrococcus lysodeikticus* was purchased from Sigma Chemical Co. Dr. J. Felsberg (Institute of Microbiology, Academy of Sciences of the Czech Republic, Prague) kindly provided *Escherichia coli* and *Bacillus subtilis*. Soluble peptidoglycan from *M. lysodeikticus* was a kind gift of Dr. J. Coyette (Department of Bacteriology, University of Liège, Belgium). Synthetic analogs of peptidoglycan fragments (muramic acid, muramyl dipeptide (MurNAc-L-Ala-D-isoglutamine, MDP), disaccharide dipeptide (GlcNAc-β-1,4-MDP), and peptidoglycan pentapeptide (L-Ala-D-isoglutaminyl-L-Lys-D-Ala-D-Ala)), LPS (*E. coli* 055:B5 S strain), β-1,3-glucans curdlan (insoluble) and laminarin (soluble), *N,N'*-diacetylchitobiose (soluble and insolubilized on 4% agarose beads), *N*-acetylglucosamine, hen egg lysozyme, and trypsin were purchased from Sigma Chemical Co. Bacteria and other reagents were suspended in 50 mM Tris-HCl, pH 8, 50 mM NaCl (TN buffer) at 1 mg/ml and sonicated until clarification. Insoluble curdlan was suspended in water (10 mg/ml) and sonicated. Further dilutions were performed in TN buffer. LPS contamination of buffers or samples other than LPS was below the detection limit or <35 pg/mg, respectively, as assessed by QCL LAL test (Bio-Whittaker, Europe).

RNA Preparations—Body pieces of adult *E. foetida* or *L. terrestris* earthworms kept for 2 days on filter paper soaked with isotonic phosphate-buffered saline in the presence of penicillin/streptomycin (100 units, 100 µg/ml) were excised and placed in liquid nitrogen, and total RNAs were prepared using TRIzol (Sigma). All molecular biology procedures were performed according to the manufacturer's recommendations.

Cloning of L. terrestris CCF—An *E. foetida* CCF (ECCF) homolog was detected in *L. terrestris* by Northern blot using total ECCF cDNA as a probe; a 1.6-kb mRNA was observed.[2] *L. terrestris* CCF (LCCF) cDNA was cloned using 5'-rapid amplification of cDNA ends (RACE) and 3'-RACE (Life Technologies, Inc.) with generated specific primers (GSP, Table I). For 5'-RACE, GSP-1 was based on ECCF and glucanase motif homologous sequences (3, 14). PCR products were gel-purified and sequenced (Thermo Sequenase radiolabeled terminator cycle sequencing kit, Amersham Pharmacia Biotech). The complete cDNA encoding LCCF was amplified (Titan One tube RT-PCR system, Roche Molecular Biochemicals) using RNA as template and primer pairs (LCCF, Table I) containing a *Bam*HI or a *Sma*I site at the 5'-N-terminal or 3'-C-terminal ends, respectively.

E. foetida and L. terrestris CCF Vector Constructions and Bacterial Expressions—The cDNA encoding *E. foetida* (ECCF, E1–E5) or *L. terrestris* (LCCF, L2, L4) CCF peptides were amplified (Titan One tube RT-PCR system) using RNA as templates and primer pairs (Table I)

---

[2] M. Bilej, P. De Baetselier, E. Van Dijck, B. Stijlemans, A. Colige, and A. Beschin, unpublished observations.

TABLE I
Primer sequences for RT-PCR reactions

		Sense	Antisense	Corresponding nucleotides in *E. foetida* CCF
5'-RACE	GSP-1	5'-CCT CGA TCG GGC GAG ATC GAC ATC ATT-3'	5'-AAT GAT GTC GAT CTC GCC CGA TCG AGG-3'	526–553
	GSP-2	5'-TGG TGG GTG GAT TTC TGG GAG-3'	5'-TTG GAA TTC GCT GTT CCC TCC GCC-3'	142–166
3'-RACE	GSP-1	5'-TTC ACC GAC TGG GAT CAA TAT C-3'		526–553
	GSP-2	5'-TTC ACT GAC TGG GAT CAG CAC C-3'		819–840
ECCF		5'-TTC ACC GAC TGG GAT CAA TAT C-3'	5'-GTT GCG CTT GTA GAC TCG-3'	51–73/1134–1152
LCCF		5'-TTC ACC GAC TGG GAT CAA TAT C-3'	5'-ATT GCG TTT GTA GAC TCG-3'	51–73/1128–1146
E1		5'-GTT GCG CTT GGA AGG CTG GTC AG-3'	5'-TTT CGG AAG GCT GGT CAG-3'	448–470/1134–1152
E2		5'-GTC CAC GCC AAA ATG CCC GTC G-3'	5'-GTT GCG CTT GTA GAC TCG-3'	448–470/1128–1146
L2		5'-TTC ACC GAC TGG GAT CAA TAT C-3'	5'-ATT GCG TTT GTA GAC TCG-3'	51–73/429–447
E3		5'-GTT CAC GCC AAG ATG CCC GTC G-3'	5'-AAC GAC AAC GCG TCC GTG GGT-3'	448–470/663–681
E4		5'-GTC CAC GCC AAA ATG CCC GTC G-3'	5'-TTT CGG AAG GCT GGT CAG-3'	448–470/663–681
L4		5'-GTT CAC GCC AAG ATG CCC GTC G-3'	5'-TTT CGG AAG ACT GGT CAG-3'	682–700/1134–1152
E5		5'-CAC TCA GAC GAT TGG AAC-3'	5'-GTT GCG CTT GTA GAC TCG-3'	

FIG. 1. **Activation of the pro-PO cascade in *E. foetida* coelomic fluid.** *A*, *M. lysodeikticus* or *E. coli* were used as such or were pretreated with coelomic fluid proteins (*CF*), lysozyme, or trypsin for the indicated times before testing their ability to trigger pro-PO activation in *E. foetida* coelomic fluid. PO activity elicited in the absence of bacteria in coelomic fluid either untreated or treated with lysozyme or trypsin is indicated by *black bars*. *, $p < 0.05$ as compared with background values. *B*, *M. lysodeikticus* peptidoglycan was used as such or was pretreated with lysozyme or CF for the indicated times before testing the pro-PO activation in *E. foetida* coelomic fluid. *Black bars* indicate PO activity in coelomic fluid in the absence of triggering agents. *, $p < 0.05$ as compared with background values.
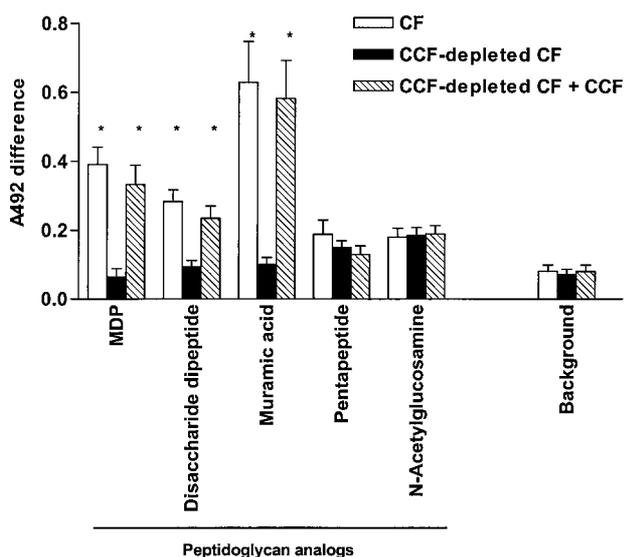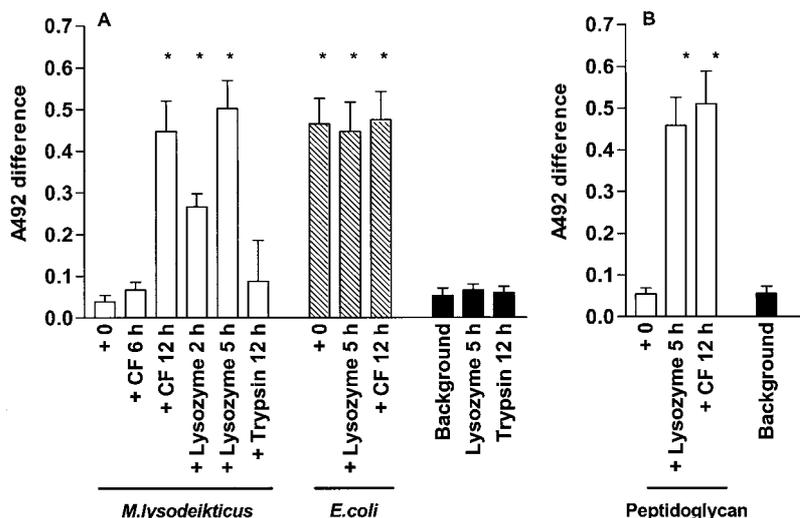
FIG. 2. **Activation of the pro-PO cascade in *E. foetida* coelomic fluid by synthetic peptidoglycan analogs.** The ability of synthetic peptidoglycan analogs (see text) to trigger *E. foetida* pro-PO activity was evaluated in coelomic fluid, CCF-depleted coelomic fluid, or in CCF-depleted coelomic fluid reconstituted with *E. foetida* CCF. PO activity elicited in the absence of peptidoglycan analogs is also indicated. *, $p < 0.05$ as compared with background values.
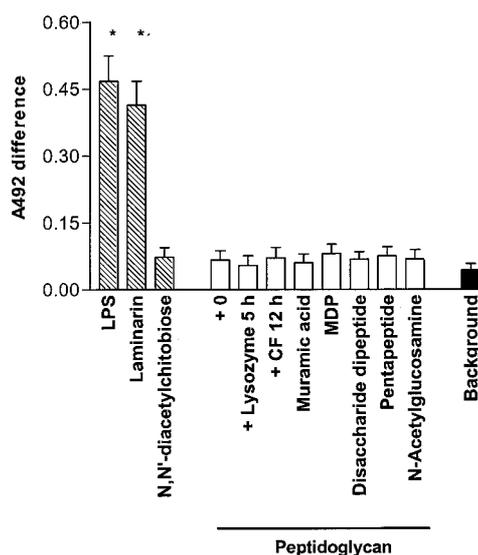
FIG. 3. **Activation of the pro-PO cascade in *L. terrestris* coelomic fluid.** PO activity triggered in *L. terrestris* coelomic fluid by LPS, laminarin, *N,N'*-diacetylchitobiose, peptidoglycan either untreated or treated with lysozyme (5 h) or coelomic fluid proteins (*CF*, 12 h), or synthetic peptidoglycan analogs was evaluated. *Black bars* indicate PO activity in coelomic fluid in the absence of triggering agents. *, $p < 0.05$ as compared with background values.

containing a *Bam*HI or a *Sma*I site at the 5′-N-terminal or 3′-C-terminal ends, respectively. After digestion, PCR products were ligated in QIAexpress PQE-30 vector containing a N-terminal $6\times$ His-affinity tag (Qiagen). The TOP10F′ *E. coli* strain was transformed, grown in the presence of ampicillin (100 μg/ml), and induced for 2 h at 37 °C in the presence of isopropyl-1-thio-β-D-galactopyranoside (2 mM). CCF peptides were purified and resuspended in TN buffer as described previously (3). LPS contamination was <15 pg/mg CCF peptides (QCL LAL test, Bio-Whittaker Europe).

Prophenoloxidase-activating Assay—The level of pro-PO cascade activation was assessed as described previously (3). Briefly, 50 μl of the coelomic fluid (without or with 1 mM Pefabloc (serine proteinase inhibitor, Roche Molecular Biochemicals)), 25 μl of 0.1 M Tris, pH 8, containing 50 mM Ca^{2+} and 10 μl of L-DOPA (3-(3,4-dihydroxylphenyl)-L-alanine, Fluka, final concentration 1.5 mM) was incubated at room temperature for 6 h in the absence or presence of tested compounds (1 μg/ml). The oxidation of L-DOPA was measured at 492 nm and expressed as the difference between the values without and with Pefabloc.

In some experiments, *M. lysodeikticus* and *E. coli* suspensions or *M. lysodeikticus* peptidoglycan were preincubated with hen egg lysozyme or trypsin (1 μg/mg bacteria or 1 μg/mg peptidoglycan, up to 5 h), or with coelomic fluid proteins (1 mg/mg bacteria or 1 mg/mg peptidoglycan, up to 12 h) before testing the activation of the pro-PO cascade.

CCF-depleted coelomic fluid was obtained by incubating the coelomic

fluid with an anti-CCF immunoaffinity column for 1 h at 4 °C (2, 3). After centrifugation, CCF-depleted coelomic fluid was used in L-DOPA oxidation test as described above. To reconstitute the pro-PO-activating cascade, equivalent concentrations of recombinant CCF or CCF peptides (25 pmol/ml) were added to CCF-depleted coelomic fluid before testing L-DOPA oxidation.

Glucan and N,N′-Diacetylchitobiose Binding Assay—CCF and CCF peptides in TN buffer (25 nmol in 500 μl) were incubated for 1 h at 4 °C with curdlan particles (10 mg/ml; 100 μl) or with agarose-insolubilized *N,N′*-diacetylchitobiose (50 μl). After washings in phosphate-buffered saline (1 ml, 13 000 × *g*, 10× for 2 min) the bound material was eluted in 50 μl of 0.0625 M Tris-HCl buffer, pH 6.8, containing 2% SDS, 20% glycerol, and 5% β-mercaptoethanol, boiled (6 min) and subjected to SDS-polyacrylamide gel electrophoresis (12%). To investigate the specificity of the interaction with curdlan or *N,N′*-diacetylchitobiose, CCF peptides were preincubated with soluble laminarin or *N,N′*-diacetylchitobiose (100 μg/100 μl; 30 min at 4 °C) before testing the ability to bind curdlan or agarose-insolubilized *N,N′*-diacetylchitobiose as described above. Coomassie Blue-stained proteins were quantified densitometrically using Analytical Imaging Station software (Imaging Research Inc., St. Catharines, Ontario, Canada).

Data Analysis—Data (mean of triplicate ± S.D.) are representative of three independent experiments performed. The validity of the results was assessed by Student's *t* test.

FIG. 4. **Amino acid sequence alignment of *E. foetida* (GenBank™ accession number AF030028) and *L. terrestris* CCF (GenBank™ accession number AF395805).** Amino acids are *numbered from the initial methionine.* Only mutated amino acids in *L. terrestris* CCF (*lt*) as compared with *E. foetida* CCF (*ef*) are indicated. Signal peptide in *L. terrestris* CCF (*underlined*) was deduced from the known signal sequence of *E. foetida* CCF (3). Polysaccharide-binding motif (*PsBM*) and glucanase motif (*GM*) are *boxed.* Tryptophan residues are in *boldface.*

```
Ef  MRWTLVVVLCL LFGEGFAFTD WDQYHIVWQD EFDYFDGAKW QHEVTATGGG NSEFQLYTQD GANSFVRDGK   70
Lt  ----A----- -LS----- -- ---------- ---F---G-- ---------- ---------- SR--------

Ef  LFIKPTLLAD NINPQTGAPF GTDFMYNGVL DVWAMYGACT NTDNNGCYRT GAAGDIPPAM SARVRTFQKY  140
Lt  -----S---- -N-----Q-Y -----N---- ----N----- ---------- --------V- ---------F

                       PsBM                                 GM
Ef  SFTHGRVVVH AKMPVGDWLW PAIWMLPEDW VYGGWPRSGE IDIIETIGNR DFKNTGGEFL GIQKMGSTMH  210
Lt  ---------- ---------- --------N- ---------- ---------- ---S-S---- ----------

Ef  WGPGWDDNRY WLTSLPKHSD DWNYGDNFHT FWFDWSPNGL RFFVDDENQA LLDVPYPLID ANPWWVDFWE  280
Lt  ---------- --------D- G--------- --Y------- ---------- ---------- ----------

Ef  WGKPWLPQYE NDNPWAGGTN LAPFDQNFHF ILNVAVGGTN GFIPDGCINR GGDPALQKPW SNGDWYNDAM  350
Lt  -----K-E-T N**-~--A-- ---------- ---------- ---------- ---------- ------VT-~  348

Ef  RKFFDARGNW KWTWDDEGDN NAMQVDYIRV YKRN                                        384
Lt  QN-Y------ ~----N---- --L------- ----                                        382
```

RESULTS

Pattern Recognition Specificity of E. foetida CCF—The coelomic fluid of *E. foetida* earthworms displays phenoloxidase (PO) activity within 6 h upon recognition of yeast or Gram-negative bacteria (3, 15). With a similar experimental procedure, Gram-positive bacteria did not trigger the coelomic fluid PO activity. However, the coelomic fluid of *E. foetida* was described to display lysozyme-like activity, with 50 µg of coelomic fluid proteins exhibiting an activity comparable with 1 µg of hen egg lysozyme in *M. lysodeikticus* bacteriolytic assay (20, 21). Therefore, we tested whether PO activity could be triggered in *E. foetida* coelomic fluid by *M. lysodeikticus* pretreated with coelomic fluid or hen egg lysozyme. As shown in Fig. 1*A*, whereas intact *M. lysodeikticus* did not trigger detectable PO activity within 6 h of incubation in coelomic fluid, bacteria pretreated for 12 h did so with considerable efficiency. In addition, bacteria pretreated for 5 h with lysozyme, but not with trypsin (up to 12 h), activated the pro-PO cascade in *E. foetida* coelomic fluid. Similar results were observed with *B.* subtilis.[2] In contrast, Gram-negative *E. coli* suspension triggered similar levels of pro-PO activity whether or not predigested with lysozyme or coelomic fluid proteins (Fig. 1*A*). On the other hand, peptidoglycan purified from the *M. lysodeikticus* cell wall did not induce PO activity unless preincubated with hen egg lysozyme for 5 h or with coelomic fluid for 12 h (Fig. 1*B*). Because lysozyme hydrolyzes the β-1,4-glycosidic bond between *N*-acetylglucosamine and *N*-acetylmuramic acid of peptidoglycan, we tested the ability of synthetic peptidoglycan analogs (muramyl dipeptide, disaccharide dipeptide, muramic acid, pentapeptide, and *N*-acetylglucosamine) to trigger the pro-PO activation in *E. foetida* coelomic fluid. Muramyl residue-containing analogs and muramic acid efficiently activated the pro-PO cascade, whereas pentapeptide and *N*-acetylglucosamine did not (Fig. 2). The PO-inducing activity of Gram-positive bacteria, or peptidoglycan digested with lysozyme or coelomic fluid proteins,[2] and of muramyl dipeptide, disaccharide dipeptide, and muramic acid (Fig. 2) was abolished in CCF-depleted coelomic fluid and reconstituted by adding recombinant CCF.

Together, these data indicate that *E. foetida* CCF displays a broad pattern recognition specificity for yeast as well as Gram-negative and Gram-positive bacteria.

Pattern Recognition Specificity of L. terrestris CCF —The coelomic fluid of another oligochaete species, *L. terrestris*, contains a 42-kDa protein cross-reacting with monoclonal antibody elicited against *E. foetida* CCF, indicating the existence of a CCF-like molecule.[2] PO activity was induced by LPS and laminarin (β-1,3-glucan) in *L. terrestris* coelomic fluid (Fig. 3). However, in contrast to *E. foetida*, neither *N,N*'-diacetylchito-
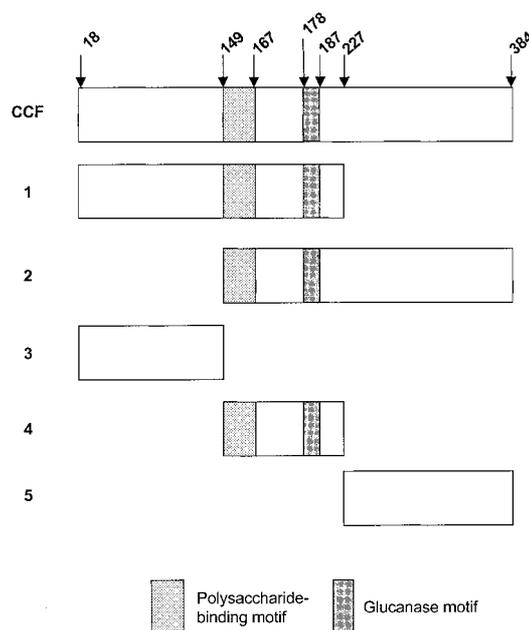

FIG. 5. **Peptide constructs generated from *E. foetida* or *L. terrestris* CCF.** Peptides covering the two putative functional regions of CCF, as well as the tryptophan-rich domain, expressed in bacteria are illustrated. The polysaccharide-binding motif and glucanase motif are indicated. *Numbers* refer to the amino acid sequences shown in Fig. 4.

biose (GlcNAc-β-1,4-GlcNAc), nor peptidoglycan hydrolyzed with lysozyme or coelomic fluid proteins, nor synthetic peptidoglycan analogs (muramic acid, muramyl dipeptide, disaccharide dipeptide, pentapeptide, or *N*-acetylglucosamine) activated the PO cascade in *L. terrestris* coelomic fluid. These results suggest that *L. terrestris* CCF has a narrower pattern recognition specificity than *E. foetida* CCF.

Comparison of the Primary Structure of E. foetida and L. terrestris CCF Molecules—*L. terrestris* CCF-like molecule was cloned, and the amino acid sequence was compared with the sequence of *E. foetida* CCF (Fig. 4). The two molecules displayed 91% of homology, with identity in the putative polysaccharide-binding motif of bacterial glucanase (7, 22) and in the putative catalytic site of bacterial endo-1,3(4)-β-glucosidase (23). Both *E. foetida* and *L. terrestris* CCF molecules contained three conserved cysteine residues and a remarkably high number of tryptophan residues (7%), the majority of which are located in the C-terminal part of the molecule.

Peptide constructs covering the two putative functional regions of CCF, as well as the tryptophan-rich domain, were expressed in bacteria to identify the carbohydrate recognition

TABLE II
Binding of CCF peptides to insoluble β-1,3-glucans or insolubilized N,N'-diacetychitobiose

E. foetida or *L. terrestris* CCF and CCF peptides, used as such or preincubated with soluble laminarin or *N,N'*-diacetylchitobiose, were incubated with curdlan (insoluble β-1,3-glucan) or with agarose-insolubilized *N,N'*-diacetylchitobiose. Bound material was eluted, subjected to SDS-polyacrylamide gel electrophoresis, and Coomassie Blue-stained. The intensity of proteins eluted without or after preincubation with laminarin or *N,N'*-diacetylchitobiose were compared to determine the percent binding inhibition.

	Binding to curdlan		Inhibition with laminarin		Binding to N,N'-diacetychitobiose		Inhibition with N,N'-diacetychitobiose	
	E. foetida	*L. terrestris*	*E. foetida*	*L. terrestris*	*E. foetida*	*L. terrestris*	*E. foetida*	*L. terrestris*
			%[a]				%[a]	
CCF	+	+	80	93	+	−	85	
Peptide 1	+	ND[b]	75		−	ND		
Peptide 2	+	+	90	98	+	−	95	
Peptide 3	−	ND			−	ND		
Peptide 4	+	+	91	95	−	−	82	
Peptide 5	−	ND			+	ND		

[a] Mean inhibition of triplicates, representative of three independent experiments; S.D. were <10% of the mean inhibition.
[b] ND, not done.

domain(s) (Fig. 5). Peptide 1 consisted of the N-terminal part of the molecule, including the polysaccharide-binding and glucanase motifs. Peptide 2 starting from the polysaccharide-binding and glucanase motifs comprised the tryptophan-rich C-terminal part as well. Peptide 3 represented the N-terminal part of CCF but did not include any of the putative functional regions. Peptide 4 corresponded to the 78 amino acids encompassing the polysaccharide-binding and glucanase motifs. Peptide 5 represented the tryptophan-rich C-terminal part of CCF. The entire molecule and all the above-mentioned peptide constructs were derived from *E. foetida* CCF (ECCF, E1, E2, E3, E4, E5). In *L. terrestris*, the entire protein and peptides 2 and 4 (LCCF, L2, L4) were used in further experiments.

Binding of CCF and CCF-derived Peptides to β-1,3-Glucans and N,N'-Diacetychitobiose—CCF and derived peptides from both *E. foetida* and *L. terrestris* were subjected to curdlan (β-1,3-glucan) or *N,N'*-diacetylchitobiose binding assay. To exclude nonspecific interactions, protein samples were preincubated either with soluble laminarin or *N,N'*-diacetylchitobiose prior to the saccharide binding assay. As shown in Table II, all tested proteins, except for peptides E3 and E5, bound specifically to β-1,3-glucans, showing the importance of polysaccharide-binding and/or glucanase motifs for interaction of CCF with β-1,3-glucans. On the other hand, only ECCF, E2, and E5, *i.e.* peptides comprising the C-terminal region of CCF, recognized *N,N'*-diacetylchitobiose. None of the *L. terrestris* CCF constructs interacted with *N,N'*-diacetylchitobiose.

Identification of CCF Pattern Recognition Domains Involved in Prophenoloxidase Activation—The binding experiments indicated that distinct domains of CCF recognized different saccharide moieties. This assumption was tested in the pro-PO activation test.

E. foetida coelomic fluid, but not CCF-depleted coelomic fluid, exhibited PO activity upon activation with LPS, laminarin, *N,N'*-diacetylchitobiose, and muramic acid (Table III). In CCF-depleted coelomic fluid, the PO activity was completely reconstituted in the presence of the four triggering agents by adding ECCF or E2. Peptides E1 and E4 containing polysaccharide-binding and glucanase motifs reconstituted the pro-PO cascade only in the presence of LPS and laminarin, whereas E5 covering the C-terminal part of ECCF restored the PO activity triggered by *N,N'*-diacetylchitobiose and muramic acid. The CCF N-terminal part E3 did not restore the PO activity assessed by any of the triggering agents in CCF-depleted coelomic fluid.

As mentioned above (Fig. 3), the pro-PO cascade was activated in the coelomic fluid of *L. terrestris* by LPS and laminarin but not by *N,N'*-diacetylchitobiose or muramic acid (Table IV). LCCF, L2, and L4 reconstituted the sensitivity to LPS and laminarin in *L. terrestris* CCF-depleted coelomic fluid. Interestingly, ECCF, E2, and E5 conferred the ability to activate the pro-PO cascade in *L. terrestris* CCF-depleted coelomic fluid by *N,N'*-diacetylchitobiose and muramic acid.

Collectively, these data suggest that CCF consists of two distinct carbohydrate recognition domains. The first (amino acids 149–227), surrounding the polysaccharide and glucanase motifs, mediates interaction of CCF with β-1,3-glucans and LPS. The second domain located in the C-terminal region CCF recognizes β-1,4-*N*-acetylglucosamine-linked saccharides and muramic acid in *E. foetida* CCF only. The reconstitution of PO activity in CCF-depleted *L. terrestris* coelomic fluid by *E. foetida* CCF peptides suggests that all components triggering the pro-PO cascade except the pattern recognition specificity of CCF are common to both earthworm species.

DISCUSSION

CCF was previously described to recognize cell wall components of yeast and Gram-negative bacteria that trigger the activation of pro-PO cascade in the coelomic fluid of *E. foetida* earthworm (3). Under similar experimental conditions, we did not detect activation of pro-PO with intact Gram-positive bacteria. Furthermore, it was reported that CCF agglutinated smooth Gram-negative bacteria, but not rough Gram-negative bacteria or Gram-positive bacteria (3). Here we show that preincubation of Gram-positive bacteria with lysozyme triggers the pro-PO cascade in *E. foetida* coelomic fluid. Lysozyme treatment does not affect the pro-PO activation by Gram-negative bacteria. The coelomic fluid of *E. foetida* earthworm exhibits a lysozyme-like activity mediated by a 13-kDa protein displaying considerable homology to other invertebrate lysozymes (20, 21). Accordingly, preincubation of Gram-positive bacteria with *E. foetida* coelomic fluid for 12 h resulted in pro-PO activation comparable with 5-h lysozyme pretreatment.

Trypsin pretreatment has no effect on the inability of Gram-positive bacteria to induce pro-PO cascade suggesting that the β-1,4-*N*-acetylmuramidase activity of lysozyme rather than proteolysis allows the recognition of Gram-positive bacteria cell wall components by earthworm defense molecule(s). Similarly, peptidoglycan, but not the synthetic peptidoglycan analogs muramyl dipeptide, disaccharide dipeptide, or muramic acid, required lysozyme pretreatment to trigger L-DOPA oxidation in *E. foetida* coelomic fluid. Moreover, synthetic pentapeptide did not exhibit significant PO-inducing activity. Hence, it can be suggested that recognition of peptidoglycan saccharide moi-

TABLE III

Reconstitution of pro-PO cascade in E. foetida CCF-depleted coelomic fluid by E. foetida CCF peptides

Pro-PO activation triggered by LPS, laminarin, N,N'-diacetylchitobiose, and muramic acid was determined in E. foetida coelomic fluid (*CF*) or in CCF-depleted coelomic fluid reconstituted with E. foetida CCF peptides (see Fig. 5). Data are shown as mean of triplicates ± S.D., representative of three independent experiments.

E. foetida	CF	Depleted CF	Depleted CF + ECCF	Depleted CF + E1	Depleted CF + E2	Depleted CF + E3	Depleted CF + E4	Depleted CF + E5
LPS	0.618 ± 0.086	0.058 ± 0.016[a]	0.426 ± 0.063	0.406 ± 0.048	0.406 ± 0.051	0.044 ± 0.015[a]	0.372 ± 0.073	0.074 ± 0.023[a]
Laminarin	0.564 ± 0.061	0.048 ± 0.011[a]	0.395 ± 0.071	0.392 ± 0.055	0.448 ± 0.073	0.059 ± 0.019[a]	0.381 ± 0.059	0.085 ± 0.031[a]
N,N'-Diacetylchitobiose	0.486 ± 0.054	0.044 ± 0.022[a]	0.420 ± 0.054	0.046 ± 0.017[a]	0.405 ± 0.063	0.052 ± 0.017[a]	0.057 ± 0.019[a]	0.379 ± 0.055
Muramic acid	0.631 ± 0.089	0.054 ± 0.021[a]	0.431 ± 0.059	0.077 ± 0.021[a]	0.448 ± 0.077	0.061 ± 0.027[a]	0.061 ± 0.021[a]	0.409 ± 0.067

[a] Lower than the CF values ($p < 0.01$).

TABLE IV

Reconstitution of pro-PO cascade in L. terrestris CCF-depleted coelomic fluid by L. terrestris or E. foetida CCF peptides

Pro-PO activation triggered by LPS, laminarin, N,N'-diacetylchitobiose, and muramic acid was determined in L. terrestris coelomic fluid (*CF*) or in CCF-depleted coelomic fluid reconstituted with L. terrestris or E. foetida CCF (see Fig. 5). Data are shown as mean of triplicates ± S.D., representative of three independent experiments.

L. terrestris	CF	Depleted CF	Depleted CF + LCCF	Depleted CF + ECCF	Depleted CF + L2	Depleted CF + E2	Depleted CF + L4	Depleted CF + E5
LPS	0.468 ± 0.088	0.073 ± 0.019[a]	0.402 ± 0.099	0.366 ± 0.057	0.461 ± 0.089	0.385 ± 0.101	0.389 ± 0.057	0.081 ± 0.019[a]
Laminarin	0.414 ± 0.073	0.069 ± 0.015[a]	0.441 ± 0.071	0.392 ± 0.084	0.421 ± 0.077	0.386 ± 0.079	0.407 ± 0.095	0.096 ± 0.025[a]
N,N'-Diacetylchitobiose	0.081 ± 0.023	0.074 ± 0.015	0.087 ± 0.019	0.355 ± 0.057[b]	0.082 ± 0.019	0.372 ± 0.073[b]	0.076 ± 0.019	0.475 ± 0.102[b]
Muramic acid	0.075 ± 0.021	0.079 ± 0.027	0.092 ± 0.023	0.348 ± 0.078[b]	0.091 ± 0.027	0.468 ± 0.103[b]	0.091 ± 0.021	0.401 ± 0.078[b]

[a] Lower than the CF values ($p < 0.01$).
[b] Higher than the CF values ($p < 0.01$).

eties rather than recognition of amino acid determinants by earthworm defense molecule(s) is required to elicit pro-PO activation. The PO-inducing activity of muramic acid, muramyl dipeptide, and disaccharide dipeptide was abolished in CCF-depleted coelomic fluid and was recovered by recombinant CCF. Hence, CCF plays a central role in the activation of the pro-PO cascade in *E. foetida* coelomic fluid, recognizing, besides yeast and Gram-negative bacteria cell wall components (3, 15), Gram-positive bacteria cell wall components, at least muramic acid.

The coelomic fluid of the related earthworm species *L. terrestris* displays distinct biochemical properties as compared with *E. foetida* coelomic fluid reflecting most likely the antigenicity of the biotope where they live (soil *versus* compost). For example, proteolytic and hemolytic activities are considerably lower in *L. terrestris* than in *E. foetida* coelomic fluid (24, 25). Moreover, in contrast to the *E. foetida* coelomic fluid (2, 15), the coelomic fluid of *L. terrestris* does not display lytic activity against murine cell lines or African trypanosomes.[2] Furthermore, here we report that the pro-PO cascade in *L. terrestris* coelomic fluid can be triggered by β-1,3-glucans and LPS but not by peptidoglycan treated with lysozyme, muramic acid, or *N,N'*-diacetylchitobiose. These results indicate that the specificity and/or function of pattern recognition molecules in *E. foetida* and *L. terrestris* differ. The comparison of the amino acid sequence of CCF from *E. foetida* and *L. terrestris* revealed 91% homology with a complete identity in the putative polysaccharide-binding motif and in the catalytic region of bacterial and animal endo-1,3(4)-β-endoglucosidases (7, 22, 23). These two regions are highly conserved in invertebrate pattern recognition molecules homologous to CCF that were reported to bind LPS or β-1,3-glucans (14). The remarkable number of tryptophan residues located mainly in the C-terminal part of CCF is of particular interest, because tryptophan residues were described to be important for the recognition of *N*-acetyl saccharides such as *N,N'*-diacetylchitobiose by tachylectin-2, a defense molecule from the horseshoe crab *T. tridentatus* (26). The crucial role of polysaccharide-binding and glucanase motifs in the recognition of LPS and β-1,3-glucans, and of the C-terminal part of CCF for the recognition of β-1,4-linked *N*-acetyl saccharides was confirmed in binding experiments and in the pro-PO activating assay using recombinant CCF-derived molecules from both *E. foetida* and *L. terrestris*. Based on these data, we propose the existence of distinct functional carbohydrate recognition domains within the CCF. The domain comprising amino acids 149–227, present in both *E. foetida* and *L. terrestris*, binds LPS and β-1,3-glucans from Gram-negative bacteria and yeast, respectively. The C-terminal domain (amino acids 227–384), showing high divergence in *E. foetida* and *L. terrestris* CCF, displays specificity for the structurally related β-1,4-linked *N*-acetyl saccharides *N,N'*-diacetylchitobiose, muramyl dipeptide, and muramic acid. Thus, this tryptophan-rich domain allows the recognition of cell wall components of yeast or Gram-positive bacteria by *E. foetida* but not by *L. terrestris* CCF. Mutations in the C-terminal part may be responsible for the inability of *L. terrestris* CCF to bind muramic acid or *N,N'*-diacetylchitobiose. The existence of two distinct carbohydrate recognition domains may extend the protection of *E. foetida* against a variety of microorganisms or facilitate the interaction of immunocytes with the pathogen and subsequent induction of cellular defense. Indeed, one domain of CCF may interact with the pathogen, while the second domain may interact with a saccharide moiety on a putative immunocyte receptor (27–30). Alternatively, the N-terminal part of CCF (amino acids 1–148) could allow the binding to a putative receptor mediating its interaction with immunocytes upon pathogen recognition.

Peptidoglycan recognition proteins have been described from insects to vertebrates, and the level of homology points to their common origin (31). Vertebrate peptidoglycan recognition proteins inhibit the growth of Gram-positive bacteria (32). In insects, peptidoglycan recognition proteins are often involved in activation of the pro-PO cascade (33, 34). Despite its analogous function, CCF does not show primary structure homology with any described peptidoglycan recognition proteins. In addition, CCF displays no amino acid homology with lysozymes, Toll-like receptors, or vertebrate CD14. Although it recognizes LPS and β-1,3-glucans through the polysaccharide-binding and glucanase motifs as other invertebrate LPS- and β-1,3-glucan-binding proteins, CCF may recognize peptidoglycan by a different mechanism from the described peptidoglycan recognition proteins. The latter bind peptidoglycan in the absence of lysozyme treatment, whereas CCF requires lysozyme-like activity to recognize peptidoglycan. CCF also differs from other peptidoglycan recognition proteins in the specificity for peptidoglycan constituents. The minimal peptidoglycan structure required to induce silkworm antibacterial defenses consists of two *N*-acetylglucosamine-*N*-acetylmuramic acid (muramyl-dipeptide) units with peptide side chains (35). Similarly, CD14 binds specifically to (muramyl-dipeptide)$_n$ as a minimal structure of peptidoglycan (36, 37). CCF recognizes soluble muramyl-dipeptide initiating innate defense reactions against Gram-positive bacteria in *E. foetida* earthworm coelomic fluid.

In conclusion, we have provided evidence for the presence of two distinct functional domains in an invertebrate defense molecule, allowing the recognition of yeast and Gram-negative and Gram-positive bacteria.

REFERENCES

1. Janeway, C. A. (1989) *Cold Spring Harb. Symp. Quant. Biol.* **54,** 1–13
2. Bilej, M., Brys, L., Beschin, A., Lucas, R., Vercauteren, E., Hanusova, R., and De Baetselier, P. (1995) *Immunol. Lett.* **45,** 123–128
3. Beschin, A., Bilej, M., Hanssens, P., Raymakers, J., Van Dijck, E., Revets, H., Brys, L., Gomez, J., De Baetselier, P., and Timmermans, M. (1998) *J. Biol. Chem.* **273,** 24948–24954
4. Yamamoto, M., Aono, R., and Horikoshi, K. (1993) *Biosci. Biotechnol. Biochem.* **57,** 1518–1525
5. Bachman, E. S., and MacClay, D. R. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 6808–6813
6. Seki, N., Muta, T., Oda, T., Iwaki, D., Kuma, K., Miyata, T., and Iwanaga, S. (1994) *J. Biol. Chem.* **269,** 1370–1374
7. Lee, W. J., Lee, J. D., Kravchenko, V. V., Ulevitch, R. J., and Brey, P. T. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 7888–7893
8. Dimopoulos, G., Richman, A., Müller, H.-M., and Kafatos, F. C. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 11508–11513
9. Shin, S. W., Park, S. S., Park, D. S., Kim, M. G., Kim, S. C., Brey, P. T., and Park, H. Y. (1998) *Insect Biochem. Mol. Biol.* **28,** 827–837
10. Kim, Y. S., Han, S. J., Ryu, J. H., Choi, K. H., Hong, Y. S., Chung, Y. H., Perrot, S., Raibaud, A., Brey, P. T., and Lee, W. J. (2000) *J. Biol. Chem.* **275,** 2071–2079
11. Lee, S. Y., Wang, R., and Söderhäll, K. (2000) *J. Biol. Chem.* **275,** 1337–1343
12. Ochiai, M., and Ashida, M. (2000) *J. Biol. Chem.* **275,** 4995–5002
13. Ma, C., and Kanost, M. M. (2000) *J. Biol. Chem.* **275,** 7505–7514
14. Bilej, M., De Baetselier, P., and Beschin, A. (2000) *Folia Microbiol.* **45,** 283–300
15. Beschin, A., Bilej, M., Brys, L., Torreele, E., Lucas, R., Magez, S., and De Baetselier, P. (1999) *Nature* **400,** 627–628
16. Smith, V. J. (1996) *Adv. Comp. Environ. Physiol.* **23,** 75–114
17. Johansson, M. W., and Söderhäll, K. (1996) *Prog. Mol. Subcell. Biol.* **15,** 46–66
18. Söderhäll, K., and Cerenius, L. (1998) *Cur. Opin. Immunol.* **10,** 23–28
19. Valembois, P., Roch, P., and Lassègues, M. (1986) in *Immunity in Invertebrates* (Brehèlin, M., ed) pp. 74–93, Springer-Verlag, Berlin, Heidelberg, New York
20. Çotuk, A., and Dales, R. P. (1984) *Comp. Biochem. Physiol.* **78A,** 469–474
21. Ito, Y., Yoshikawa, A., Hotani, T., Fukuda, S., Sugimura, K., and Imoto, T. (1999) *Eur. J. Biochem.* **259,** 456–461
22. Yahata, N., Watanabe, T., Nakamura, Y., Yamamoto, Y., Kamimiya, S., and Tanaka, H. (1990) *Gene* **86,** 113–117
23. Juncosa, M., Pons, J., Dot, T., Querol, E., and Planas, A. (1994) *J. Biol. Chem.* **269,** 14530–14535

24. Tuckova, L., Rejnek, J., Sima, P., and Ondrejova, R. (1986) *Dev. Comp. Immunol.* **10,** 181–189
25. Leipner, C., Tuckova, L., Rejnek, J., and Langner, J. (1993) *Comp. Biochem. Physiol.* **105B,** 637–641
26. Beisel, H. G., Kawabata, S., Iwanaga, S., Huber, R., and Bode, W. (1999) *EMBO J.* **18,** 2313–2322
27. Bilej, M., Rossmann, P., Sinkora, M., Hanusova, R., Beschin, A., Raes, G., and De Baetselier, P. (1998) *Immunol. Lett.* **60,** 23–29
28. Vasta, G. R., Quesenberry, M., Ahmed, H., and O'Leary, N. (1999) *Dev. Comp. Immunol.* **23,** 401–420
29. Koizumi, N., Imamura, M., Kadotani, T., Yaoi, K., Iwahana, H., and Sato, R. (1999) *FEBS Lett.* **443,** 139–143
30. Bianchet, M. A., Ahmed, H., Vasta, G. R., and Amzel, L. M. (2000) *Proteins* **40,** 378–388
31. Kang, D., Liu, G., Lundstrom, A., Gelius, E., and Steiner, H. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 10078–10082
32. Liu, C., Gelius, E., Liu, G., Steiner, H., and Dziarski, R. (2000) *J. Biol. Chem.* **275,** 24490–24499
33. Yoshida, H., Kinoshita, K., and Ashida, M. (1996) *J. Biol. Chem.* **271,** 13854–13860
34. Ochiai, M., and Ashida, M. (1999) *J. Biol. Chem.* **274,** 11854–11858
35. Iketani, M., Nishimura, H., Akayama, K., Yamano, Y., and Morishima, I. (1999) *Insect Biochem. Mol. Biol.* **29,** 19–24
36. Gupta, D., Kirkland, T. N., Viriyakosol, S., and Dziarski, R. (1996) *J. Biol. Chem.* **271,** 23310–23316
37. Dziarski, R., Tapping, R. I., and Tobias, P. S. (1998) *J. Biol. Chem.* **273,** 8680–8690