doi:10.1016/j.jmb.2005.01.070 J. Mol. Biol. (2005) 347, 683–691

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

COMMUNICATION

Crystal Structure of Human Peptidoglycan Recognition Protein S (PGRP-S) at 1.70 Å Resolution

Rongjin Guan[1]†, Qian Wang[1]†, Eric J. Sundberg[2] and Roy A. Mariuzza[1]*

[1]*Center for Advanced Research in Biotechnology, W. M. Keck Laboratory for Structural Biology, University of Maryland Biotechnology Institute, Rockville, MD 20850 USA*

[2]*Boston Biomedical Research Institute, Watertown, MA 02472, USA*

**Corresponding author*

Peptidoglycan recognition proteins (PGRPs) are pattern recognition receptors of the innate immune system that bind peptidoglycans (PGNs) of bacterial cell walls. These molecules, which are highly conserved from insects to mammals, contribute to host defense against infections by both Gram-positive and Gram-negative bacteria. Here, we present the crystal structure of human PGRP-S at 1.70 Å resolution. The overall structure of PGRP-S, which participates in intracellular killing of Gram-positive bacteria, is similar to that of other PGRPs, including *Drosophila* PGRP-LB and PGRP-SA and human PGRP-Iα. However, comparison with these PGRPs reveals important differences in both the PGN-binding site and a groove formed by the PGRP-specific segment on the opposite face of the molecule. This groove, which may constitute a binding site for effector or signaling proteins, is less hydrophobic and deeper in PGRP-S than in PGRP-IαC, whose PGRP-specific segments vary considerably in amino acid sequence. By docking a PGN ligand into the PGN-binding cleft of PGRP-S based on the known structure of a PGRP-Iα–PGN complex, we identified potential PGN-binding residues in PGRP-S. Differences in PGN-contacting residues and interactions suggest that, although PGRPs may engage PGNs in a similar mode, structural differences exist that likely regulate the affinity and fine specificity of PGN recognition.

Keywords: crystal structure; PGRP-S; peptidoglycan recognition protein; innate immunity; pattern recognition receptor

Peptidoglycan recognition proteins (PGRPs) are a family of pattern recognition receptors (PRRs) of the innate immune system that bind and, in some cases, hydrolyze bacterial peptidoglycans (PGNs).[1–9] PGNs belong to a group of conserved microbial motifs (pathogen-associated molecular patterns, or PAMPs) that are unique products of microbial metabolism not produced by the host. Other examples of PAMPs include lipopolysaccharide, flagellin, and unmethylated CpG dinucleotides, which are recognized by Toll-like receptors, and mannans, which are bound by collectins.[1,10]

PGNs, an essential cell wall component of virtually all bacteria, are polymers of alternating *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) in β(1→4) linkage, cross-linked by short peptide stems.[11,12] Whereas the glycan chains display little variation in different bacterial species, the cross-linking peptides, which are composed of alternating L and D-amino acids, vary in length and amino acid composition. Most notably, PGNs are divided into two major categories based on the residue at position 3 of the peptide stem: L-lysine type (Lys-type) PGN from Gram-positive bacteria and *meso*-diaminopimelic acid-type (Dap-type) PGN from Gram-negative bacteria.[11,12] A number of PRRs have been shown to interact with PGNs, including nucleotide-binding oligomerization domain-containing proteins (NODs) and PGRPs.[2,13,14]

PGRPs are highly conserved from insects to mammals and show distinct specificities for PGNs from different microbes.[3] To date, more than 40 PGRPs have been identified.[12] PGRPs, which are structurally related to bacteriophage T7 lysozyme,[15] have been divided into three classes: short PGRPs

† R.G. and Q.W. contributed equally to this work.

Abbreviations used: PGRPs, peptidoglycan recognition proteins; PGNs, peptidoglycans; r.m.s., root-mean-square; MTP, muramyl tripeptide.

E-mail address of the corresponding author: mariuzza@carb.nist.gov

(PGRP-S) of 20–25 kDa, intermediate PGRPs (PGRP-I) of 40–45 kDa, and long PGRPs (PGRP-L) of up to 90 kDa. Most PGRPs are soluble intracellular or secreted proteins; so far only *Drosophila* PGRP-LC has been definitely identified as a cell surface receptor.[16] *Drosophila* has 13 PGRP genes[5] and the mosquito *Anopheles* seven.[17] Humans and other mammals encode four PGRPs: one each of the L (PGRP-L) and S (PGRP-S) forms, and two of the I (PGRP-Iα and PGRP-Iβ) form.[3,7]

Drosophila PGRPs activate two distinct signaling pathways leading to production of antimicrobial peptides. PGRP-SA recognizes Lys-type PGN, which triggers the Toll receptor pathway;[1,18] PGRP-LC and PGRP-LE interact with Dap-type PGN and activate the Imd pathway.[16,19–24] The functions of mammalian PGRPs are less well understood. Human and mouse PGRP-L, like several *Drosophila* PGRPs,[8,25] hydrolyze the amide bond between the MurNAc and L-alanine moieties of PGNs and probably act as scavengers.[9,26] Mouse PGRP-S inhibits the growth of certain Gram-positive bacteria in culture media and participates in the intracellular killing of bacteria in neutrophils,

Table 1. Data collection and structural refinement statistics

Data collection	
Resolution (Å)	1.65
Space group	$P3_221$
Cell parameters (Å)	$a=b=72.12$, $c=51.79$
Unique reflections[a]	19,042 (1090)
Completeness (%)[a]	99.8 (100.0)
R_{merge} [a,b]	0.034 (0.246)
Redundancy	6.7 (6.5)
Mean $I/\sigma(I)$	28.3 (6.4)
Refinement	
Resolution range (Å)	40.0–1.70
R_{cryst} (%)[c]	18.3
R_{free} (%)[c]	21.0
Number of reflections used	17,428
Number of reflections in R_{free} set	799
Number of non-H protein atoms	1322
Number of water molecules	200
Average temperature factors (Å^2)	
Main chain atoms	18.96
Side-chain atoms	21.51
Waters	35.93
r.m.s. derivation bond lengths (Å)	0.010
r.m.s. derivation bond angles (deg.)	1.342
Ramachandran plot statistics	
Most favored (%)	87.9
Additionally allowed (%)	10.7
Generously allowed (%)	1.4

A DNA fragment encoding residues 1–175 of human PGRP-S was generated by PCR from the corresponding cDNA (Genecopoeia) and cloned into the bacterial expression vector pT7-7 (Novagen). The protein was expressed as inclusion bodies in *Escherichia coli* BL21 (DE3) cells (Stratagene). Bacteria were grown at 37 °C to an absorbance of 0.6 at 600 nm, and isopropyl-β-D-thiogalactoside was added to a concentration of 1 mM. After incubation for three to four hours, the bacteria were harvested by centrifugation and resuspended in 50 mM Tris–HCl (pH 8.0) containing 2 mM EDTA and 0.5% Triton X-100; cells were disrupted by sonication. The inclusion bodies were washed three times with 50 mM Tris–HCl (pH 8.0), 2 mM EDTA,0.5% (w/v) Triton X-100, and another three times with the same buffer without Triton X-100, then solubilized in 50 mM Tris–HCl (pH 8.0), 8 M urea, 2 mM EDTA, 1 mM DTT. For *in vitro* folding, solubilized PGRP-S inclusion bodies were diluted into 1.0 M arginine, 50 mM Tris–HCl (pH 8.5), 2 mM EDTA, 5 mM cysteamine, 0.5 mM cystamine. After three days at 4 °C, the folding mixture was concentrated, dialyzed against 50 mM Mes (pH 5.8), and applied to a Mono S cation-exchange column (Amersham Biosciences) equilibrated in the same buffer; the protein was eluted with a linear NaCl gradient. Mass spectrometry and N-terminal sequencing established the identity of the recombinant protein. Crystallization trials of purified PGRP-S were carried out in hanging drops at a protein concentration of 6 mg/ml. Crystals of PGRP-S grew at room temperature in 30% (w/v) polyethylene glycol (PEG) 4000, 0.1 M sodium citrate (pH 5.6), 0.2 M ammonium acetate. For data collection, crystals of PGRP-S were cryoprotected by brief soaking in mother liquor containing 15% (v/v) glycerol prior to flash-cooling in a nitrogen stream. X-ray diffraction data to 1.65 Å resolution were recorded at 100 K using an R-axis IV^{++} image plate detector equipped with Osmic mirrors and mounted on a Rigaku rotating anode CuK_α X-ray generator. The data were processed and scaled with CrystalClear.[42] The structure of PGRP-S was determined by molecular replacement with the program AMoRe[43] using human PGRP-IαC (PDB accession code 1SK3) as the search probe. The translation function search gave a clear solution with a correlation coefficient of 0.291 and R_{factor} of 50.2% at a resolution range of 8.0–3.5 Å. After rigid-body refinement, the correlation coefficient was 0.481 and the R_{factor} was 47.1%. The solution contains one monomer in the asymmetric unit. Refinement of the PGRP-S model was performed with CNS version 1.1[38] with data between 40.0 Å and 1.70 Å; σ_A-weighted $2F_o-F_c$ and F_o-F_c electron density maps were calculated for model adjustment. After several rounds of rebuilding in XtalView,[44] including gradual residue replacement to the human PGRP-S sequence, R_{cryst} was reduced to 29.2% and R_{free} to 31.1%. Group temperature factor (B) refinement was done with further model adjustment, yielding R_{cryst} and R_{free} of 24.1% and 27.1%, respectively. Water molecules were added automatically with ARP/wARP,[45] followed by refinement with Refmac5[46] for isotropic B factors. The final model comprises residues 9–175 and 200 water molecules. The final R_{cryst} is 18.3% and R_{free} is 21.0% for all data between 40.0 and 1.70 Å.

[a] Values in the parentheses are statistics of the highest resolution shell (1.71–1.65 Å).

[b] $R_{merge}=\Sigma|I_j-\langle I\rangle|/\Sigma I_j$, where I_j is the intensity of an individual reflection and $\langle I\rangle$ is the average intensity of that reflection.

[c] $R_{cryst}=\Sigma||F_0|-|F_c||/\Sigma|F_0|$, where F_c is the calculated structure factor. R_{free} is as for R_{cryst} but calculated for a randomly selected 5.0% of reflections not included in the refinement.

where it is found in tertiary granules.[6] Accordingly, PGRP-S knockout mice exhibit increased susceptibility to intraperitoneal infections with Gram-positive bacteria.[27] Bovine PGRP-S, present in neutrophil and eosinophil granules, is bacteriostatic or bacteriocidal for both Gram-positive and Gram-negative bacteria.[28] In addition, mouse PGRP-S has been reported to form a potent cytotoxic complex with heat shock protein 70 (Hsp70) that induces apoptotic death in various tumor lines.[29]

The crystal structures of three PGRPs have been reported: *Drosophila* PGRP-LB[25] and PGRP-SA,[30,31] and the C-terminal PGN-binding domain of human PGRP-Iα (designated PGRP-IαC).[32] These proteins adopt a similar overall fold, including a conserved PGN-binding cleft, but differ in their N-terminal regions. Recently, we described the structure of PGRP-IαC in complex with MurNAc-L-Ala-D-isoGln-L-Lys, a muramyl tripeptide (MTP) representing the core of Lys-type PGN.[33] To obtain further insights into the relationship between insect and mammalian PGRPs, as well as the basis for PGN recognition, we determined the crystal structure of human PGRP-S at 1.70 Å resolution, the first of a mammalian PGRP belonging to the S class.

Overview of the PGSP-S structure

Recombinant human PGRP-S (residues 1–175, excluding the signal peptide) was expressed in bacterial inclusion bodies and folded *in vitro* (Table 1). The crystal structure was determined at 1.70 Å resolution by molecular replacement methods using the PGRP-IαC structure as a search model (Table 1). The final refined model consists of residues 9–175 of the PGRP domain, with no observable electron density for residues 1–8. This is also the case for *Drosophila* PGRP-LB[25] and PGRP-SA,[30,31] suggesting a high degree of flexibility for the N terminus of at least some PGRPs. The root-mean-square (r.m.s.) differences in α-carbon positions, excluding the disordered N-terminal residues, are 0.91 Å, 1.02 Å and 1.31 Å for comparisons of PGRP-S with PGRP-IαC, PGRP-SA and PGRP-LB, respectively.

The human PGRP-S structure contains a central β-sheet composed of five β-strands, four parallel and one (β5) antiparallel, and three α-helices (Figure 1). Like other PGRPs, PGRP-S also contains an N-terminal segment comprising residues 9–30 (the PGRP-specific segment), whose structure varies substantially among PGRPs (see below). PGRP-S includes three disulfide bonds formed by cysteine residues at positions 9 and 133, 25 and 70, and 46 and 52. The former two disulfides tether the PGRP-specific segment to helix α2 and loop α1-β4, respectively, and the third links the β3-α1 loop tightly to helix α1. All of these disulfide bonds are also present in human PGRP-IαC.[32] In fact, these six cysteine residues are strictly conserved in all mammalian PGRP-S proteins, as well as the C-terminal PGN-binding domains of PGRP-I molecules. One disulfide bond, Cys46–Cys52, appears to be present in all known PRGPs, with *Drosophila* PGRP-LE being the sole exception, indicating its importance in maintaining the structural integrity of the PGRP domain. The Cys9–Cys133 disulfide is found in several insect PGRP-S proteins, including *Drosophila* PGRP-SA, but in no PGRP-L. By contrast, the Cys25–Cys70 disulfide is unique to mammalian PGRPs, where it is retained in all non-catalytic PGRPs, but is absent

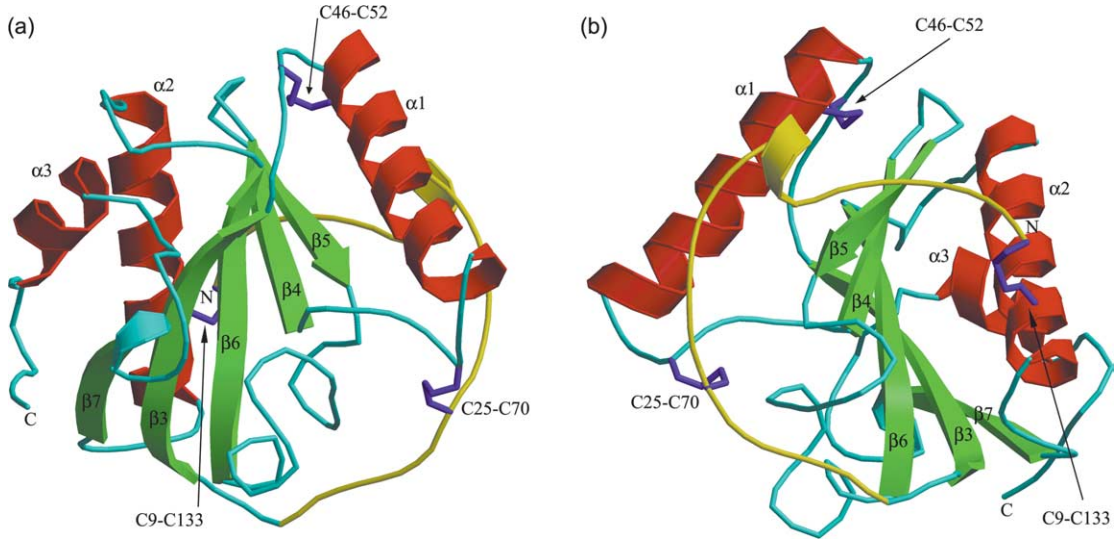

Figure 1. Ribbon diagrams of the crystal structure of human PGRP-S. (a) The view is looking down on the PGN-binding cleft, whose walls are formed by helix α1 and four loops (β3–α1, α1–β4, β5–β6 and β7–α3) extending above the β-sheet platform. Secondary structure elements are labeled following the numbering for human PGRP-Iα.[32] The N and C termini are indicated. The N-terminal PGRP-specific segment is shown in yellow. Disulfide bonds are purple. (b) View of the opposite face of the PGRP domain from that in (a), looking down onto the PGRP-specific segment. The molecule was rotated 220° about a vertical axis and 20° about a horizontal axis compared to the orientation in (a). The Figures were prepared with MOLSCRIPT[36] and rendered by Raster3D.[37]

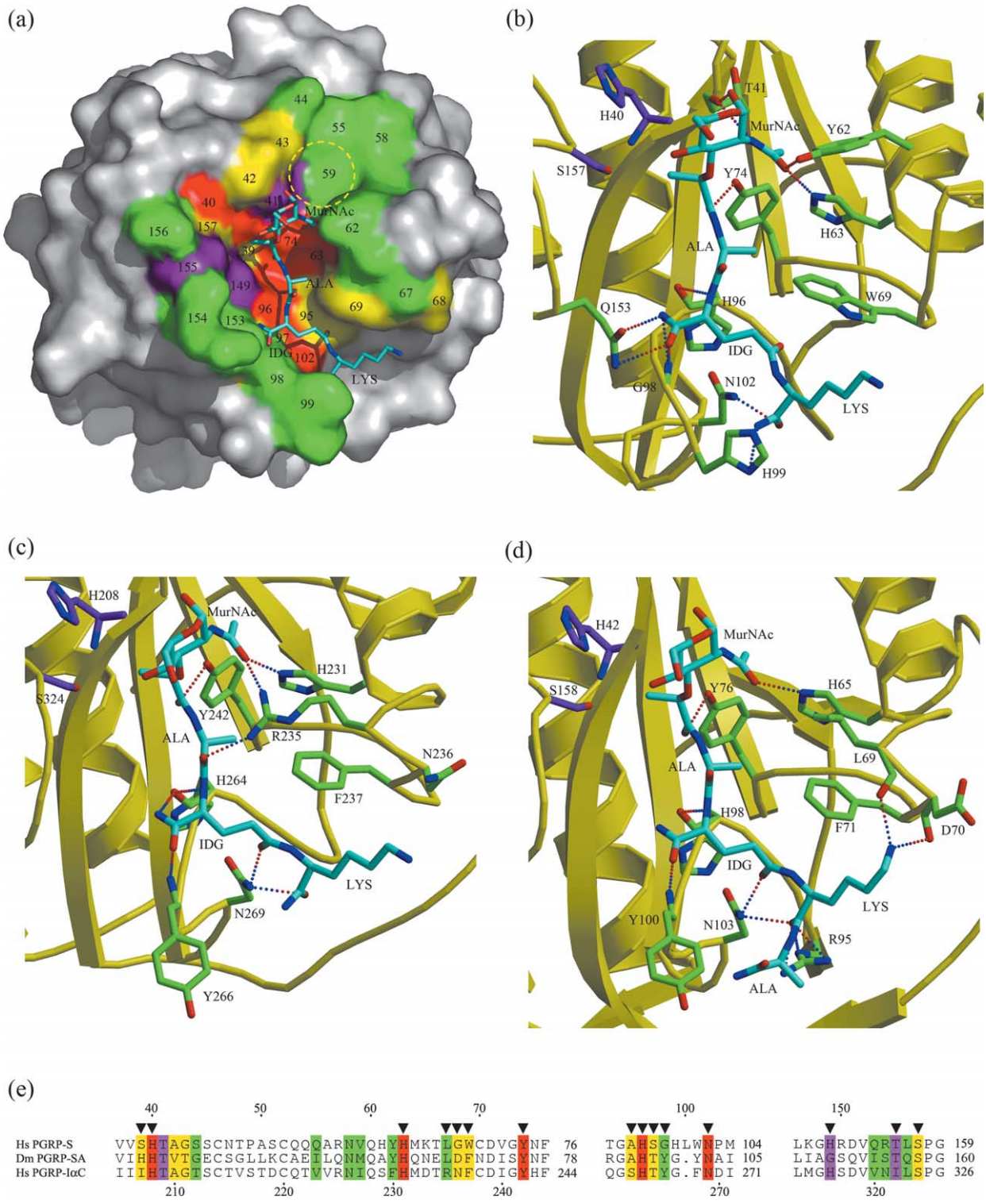

Figure 2. PGN-binding site and possible PGN-contacting residues of human PGRP-S. (a) Surface representation of the PGN-binding cleft of PGRP-S. The molecular surface is colored according to percentage identities of residues lining the PGN-binding cleft of PGRPs, based on alignments of 45 insect and mammalian sequences:[33] red, >80% identical; purple, 60–80%; yellow, 40–60%; and green, <40%. The PGN analog MTP, which was docked into the PGN binding site of PGRP-S using the PRGP-IαC–MTP structure, is shown in ball-and-stick representation. Carbon atoms are cyan, nitrogen atoms dark blue, and oxygen atoms red. A putative binding pocket for the GlcNAc moiety of natural PGNs, not present in the MTP fragment, is circled in yellow. MTP was docked into PGRP-S by superposing PGRP-IαC onto PGRP-S and then manually positioning MTP into the binding cleft of the latter; no steric clashes were observed between MTP and binding site residues. The manually docked PGRP-S–MTP model was subjected to 200 cycles of energy minimization in CNS version 1.1.[38] During minimization, the MTP ligand was free to move, MTP-contacting residues were restrained, and all other residues were fixed. The Figure was prepared with PyMOL (http://www.pymol.org). ALA, L-alanine; IDG, D-isoglutamine; LYS, L-lysine. (b) Potential interactions between PGRP-S and MTP in the docked model. MTP is cyan,

from PGRPs with known PGN-hydrolyzing activity, such as mouse and human PGRP-L.[9,26]

The conserved PGRP protein fold provides two sites for molecular interactions, one proven[33] and the other proposed.[25,30–32] The first of these is the PGN-binding site, responsible for interactions with, and in some cases a site of catalysis for, PGN moieties. This binding site resides in a long (~25 Å) cleft whose walls are formed by helix α1 and five loops (β3–α1, α1–β4, β5–β6, β6–α2 and β7–α3) that project above the β-sheet platform. The second site, located on the opposite side of the protein to the PGN-binding site, potentially accommodates host effector or signaling molecules. It is formed by the variable PGRP-specific segment and helix α2 (Figure 1(b)).

Conserved PGN-binding site

The previously reported crystal structures of *Drosophila* PGRP-LB[25] and PGRP-SA[30,31] and human PGRP-IαC[32] have revealed that all PGRPs share a conserved PGN-binding groove. PGRP-S also presents an L-shaped cleft for PGN binding (Figure 2(a)). The groove is predominantly hydrophilic and filled with numerous water molecules, which in the PRGP-IαC–MTP complex are displaced upon ligand binding, such that the interface is devoid of buried water molecules.[33] The PGN-binding cleft in PGRP-S is comprised of two distinct regions, a shallow end flanked by helix α1 and loops β3–α1 and β6–α2, while the remainder is deeper, flanked by loops α1–β4, β5–β6 and β7–α3. In the PGRP-IαC–MTP complex, the deep part of the groove accommodates the MurNAc moiety and peptide stem, and thus the shallow end likely binds GlcNAc, which links the MurNAc in natural PGNs.

Figure 2(a) shows the PGN-binding groove of PGRP-S, with the molecular surface colored based on the relative identities of residues lining the groove amongst all known PGRP sequences. Several residues that contact MTP in the PRGP-IαC–MTP complex are highly identical (>80%). Potential GlcNAc-contacting residues, however, have significantly lower identities (<40%). It has been shown that MTP is the minimum fragment of PGNs recognizable by human PGRP-L[9] as well as by PGRP-IαC.[33] This explains the ability of a wide range of related PGRP molecules to specifically bind different PGNs that present a common molecular core coupled to heterogeneous branching elements. In contrast to *Drosophila* PGRP-LB,[30] a catalytic PGRP, no zinc ion is visible in the PGN-binding site of human PGRP-S, which binds but does not hydrolyze PGNs,[9] due to replacement of zinc-coordinating residue Cys160 in PGRP-LB by Ser158 in PGRP-S. In catalytic PGRPs such as PGRP-LB, the bound Zn^{2+} acts as electrophilic catalyst to promote cleavage of the amide bond between the MurNAc and L-Ala moieties of PGNs.[33]

Possible PGN-binding residues in human PGRP-S

We have found that PGRP-S binds muramyl peptide derivatives of PGN with affinities similar to PGRP-Iα (to be published elsewhere). Although we were able to obtain the crystal structure of the PRGP-IαC–MTP complex using ligand soaking techniques, we have been unsuccessful to date in producing crystals of a PGRP-S–PGN complex by soaking or co-crystallization. In lieu of such a structure, and in order to better understand PGN recognition by human PGRP-S, we performed homology-assisted docking of MTP into the PGN-binding site of PGRP-S, as we had done previously for *Drosophila* PGRP-LB.[33] Superposition of these two structures gave an r.m.s. deviation of 0.78 Å for 140 α-carbon atoms, excluding the structurally variable PGRP-specific segment, which is opposite the PGN-binding cleft and therefore not related to PGN binding.

In the docked PGRP-S–MTP complex (Figure 2(b)), MTP forms a number of hydrogen bonds with PGRP-S, some of which are retained in the PGRP-IαC–MTP structure (Figure 2(c)), while others represent additional interactions with MTP. Aside from key residues that form conserved interactions with MTP (His40, His63, Tyr74, His96, Asn102), several PGRP-S residues make potentially important interactions with MTP that are not present in PGRP-IαC–MTP complex (Thr41, His99, Gln153) and *vice versa*. In particular, PGRP-S makes more interactions with the glycan moiety than does PGRP-IαC, with three additional hydrogen bonds *via* the main chain of Thr41 and side-chain of Tyr62 to MurNAc. In the unbound PGRP-S structure, the side-chain of His99 displays a dual conformation. However, it is probable that its conformation is fixed upon binding to PGN, due to formation of two

PGRP-S is yellow, and contacting residues are green or purple. The orientation is the same as in (a). Hydrogen bonds are drawn as broken lines; residues predicted to form multiple van der Waals contacts are also shown. (c) Interactions between PGRP-IαC and MTP in the PGRP-IαC–MTP crystal structure.[33] (d) Potential interactions between *Drosophila* PGRP-SA and MTeP (MurNAc-L-Ala-D-isoGln-L-Lys-D-Ala). The PGN ligand was docked into PGRP-SA as described for (a). In purple are predicted MTeP-contacting residues Ser158 and His42, whose mutation to alanine abolishes the L,D-carboxypeptidase activity of PGRP-SA.[31] (e) Structure-based alignment of residues lining the PGN-binding grooves of human PGRP-S, human PGRP-IαC and *Drosophila* PGRP-SA.[30–32] Binding site residues are highlighted with different colors according to their percentage identity as described in (a) (red, >80%; purple, 60–80%; yellow, 40–60%, and green, <40%); flanking sequences, included for reference only, are not colored. Triangles above the human PGRP-S sequence mark MTP-contacting residues in the PGRP-IαC–MTP crystal structure.[33]

hydrogen bonds between its side-chain and the MTP N2 atom of L-Lys in the model complex. The side-chain of Gln153 in PGRP-S interacts with MTP *via* two salt-bridges, with atoms O and N2 of D-isoGln, which also form hydrogen bonds with the main-chain nitrogen atom of Gly98. While Arg235 forms two hydrogen bonds in the PGRP-IαC–MTP complex, a leucine replacement at the analogous position 67 in PGRP-S results in the loss of these interactions in the PGRP-S–MTP model complex.

The predicted differences in the PGN-contacting residues and interactions indicate that, although PGRPs may recognize and bind PGNs in a similar mode, structural differences exist that may modulate the affinity and specificity of PGN binding. In the PGRP-IαC–MTP structure (Figure 2(c)), Asn236 and Phe237 form a number of van der Waals contacts with the side-chain of L-Lys, suggesting that sequence variability at these two positions may account for the ability of some PGRPs to discriminate between Lys-type and Dap-type PGNs. In PGRP-S, which preferentially recognizes Lys-type PGNs in binding assays using whole bacteria or insoluble PGN,[6] the corresponding sequence is Gly68-Trp69, of which only Trp69 contacts L-Lys in the PGRP-S–MTP model (Figure 2(b)). However, as significant cross-reactivity with Gram-negative bacteria was also observed,[6] the Gly68-Trp69 sequence of PGRP-S only partially restricts the PGN-binding specificity of this PGRP.

Recently, an intrinsic L,D-carboxypeptidase activity for Dap-type PGN fragments was reported for *Drosophila* PGRP-SA, involving specific cleavage of the peptide bond between *meso*-diaminopimelic acid at position 3 and D-Ala at position 4 of the peptide stem.[31] Futhermore, mutation of His42 or Ser158 to alanine abolished this hydrolytic activity. By contrast, no cleavage of Lys-type PGN fragments was observed, despite strong binding. Although the structure of a PGRP-SA–PGN complex has not been reported, we docked a muramy tetrapeptide (MTeP) ligand into the PGN-binding site of PGRP-SA based on its close similarity to human PGRP-IαC (r.m.s. difference of 0.83 Å for 140 α-carbon atoms) and on the finding that mutation of several PGRP-SA residues corresponding to PGN-contacting residues in the PGRP-IαC–MTP structure abolished recognition of Lys-type PGN.[31] In the docked PGRP-SA–MTeP complex (Figure 2(d)), His42 and Ser158 are predicted to contact the MurNAc moiety, but at the opposite end of the PGN-binding groove from residues 3 and 4 of the peptide stem. If so, the deleterious effects of mutating His42 and Ser158 on the L,D-carboxypeptidase activity of *Drosophila* PGRP-SA likely arise through alterations in the stability of the putative enzyme–substrate complex. However, in the absence of a PGRP-SA–ligand structure, we cannot formally exclude that PGRP-SA binds Dap-type PGN very differently from the way PGRP-IαC binds Lys-type PGN, possibly permitting His42 and Ser158 to participate directly in the catalytic mechanism.

The variable PGPR-specific segment

Besides their role as PGN scavengers (in the case of PGRPs with Zn^{2+}-dependent amidase activity),[8,9] PGRPs act as adaptors that link PGN recognition to intracellular signaling or complement activation.[1] To carry out this function, it has been suggested that PGRPs have evolved two structurally distinct binding sites: a highly conserved site for recognizing PGNs and a much more variable site, not involved in PGN binding, for interacting with diverse signaling or effector molecules.[25,30–32] Thus, the *Drosophila* analog of human PGRP-S, PGRP-SA, has been shown to associate with Gram-negative binding protein 1 (GNBP-1) in activating the Toll receptor pathway.[34,35] It has also been reported that mouse PGRP-S can form a stable cytotoxic complex with Hsp70 in solution and in lymphocytes.[29] All PGRPs contain an N-terminal segment of 22–33 residues (the PGRP-specific segment) characterized by high sequence variability that, with helix α2, forms a hydrophobic groove on the face opposite the PGN-binding site of each protein, and which may constitute an interaction site for non-PGN ligands, such as GNBP-1, Hsp70, or complement components.

The PGRP-specific segment of human PGRP-S consists of residues 1–30, of which only residues 9–30 are visible in the crystal structure, with a conserved central region at positions 15–21, flanked by two variable regions at positions 9–14 and 22–30 (Figure 3(a)). In the central region, residues 15–18 form a 3_{10} helix, found in all PGRP structures known to date, that links two stretches of coil in PGRP-S (Figure 1(b)). Within this helix, Arg15 and Trp18 form extensive interactions with the main body of the PGRP-S domain: Arg15 makes a bidentate salt-bridge with Asp71, while Trp18 forms a tight hydrophobic cluster with Ala20, Ala56 and Tyr85 (not shown). Trp18 is invariant among mammalian and insect PGRPs; Arg15 and Ala20 are highly conserved. Previously, we reported on differences in the PGRP-specific segment between human PGRP-IαC and *Drosophila* PGRP-LB.[32] Figure 3(b) shows a superposition of the PGRP-specific segments of PGRP-S and PGRP-IαC, which differ at 12 of 15 positions in the regions flanking the central region; PGRP-S also contains an insertion at position 28. It is clear that the two segments differ markedly in both residue type (polar side-chains or non-polar side-chains) and relative orientations of their side-chains. The result is that the groove formed by the PGRP-specific segment of PGRP-S, while still predominantly hydrophobic, is notably less so than the corresponding groove of PGRP-IαC (Figure 3(c) and (d)). In addition, this groove is deeper in PGRP-S than in PGRP-IαC. The groove formed by the PGRP-specific segment of *Drosophila* PGRP-SA also differs substantially from that of human PGRP-S (not shown), even though both are S form PRGPs. Thus, S form PGRPs from insects and

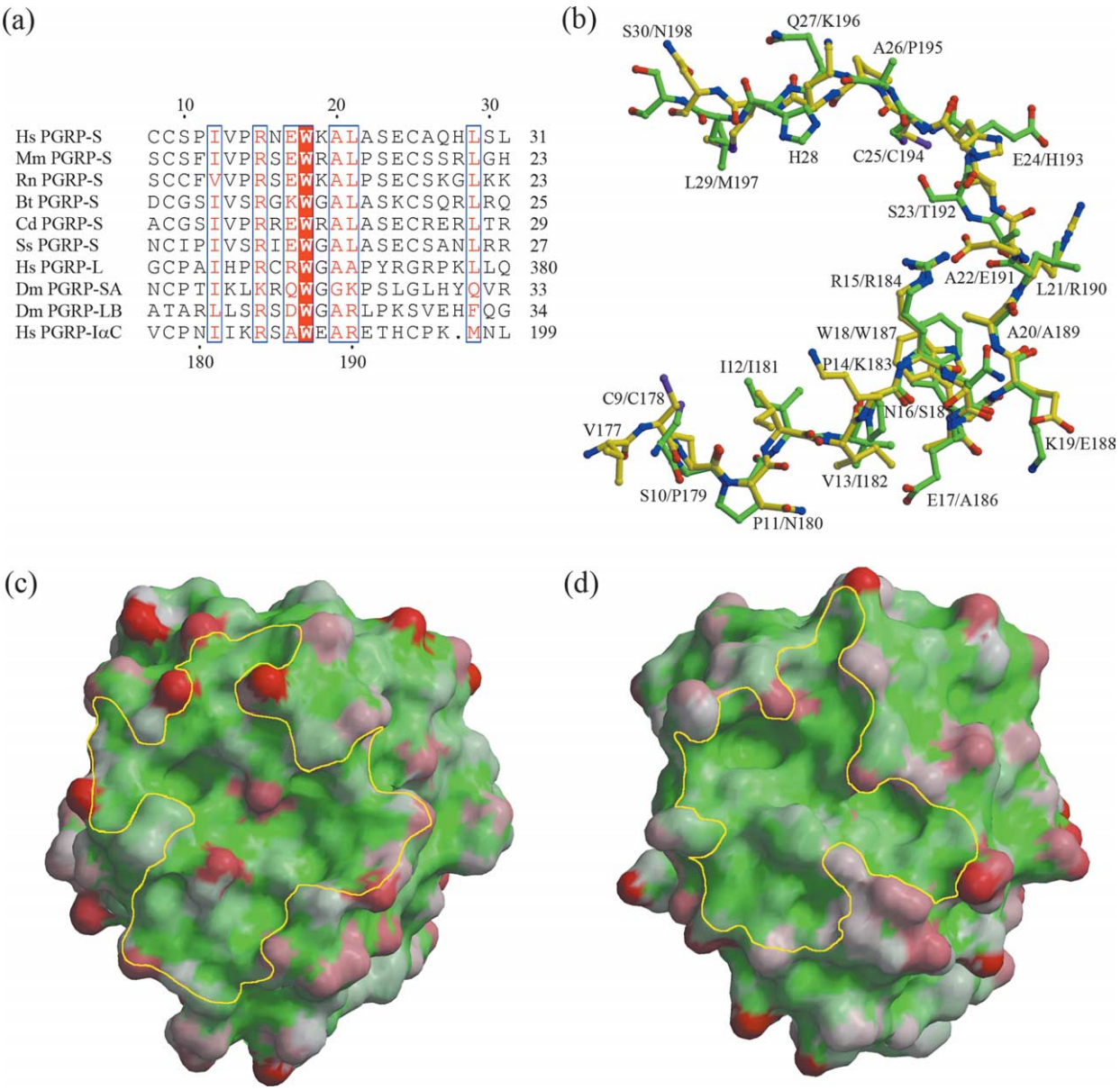

Figure 3. Structural variability in PGRP-specific segments. (a) Structure-based sequence alignment of mammalian and insect PGRP-specific segments. White characters on a red background show strictly conserved residues. Residues that are well conserved are drawn in red and framed in blue. The remaining residues are black. Abbreviations and GenBank or SwissProt accession numbers are as follows. Hs, *Homo sapiens* (PGRP-Iα, AY035376; PGRP-S, AF076483; PGRP-L, AF384856); Mm, *Mus musculus* (mouse; PGRP-S, AF076482); Rn, *Rattus norvegicus* (rat; PGRP-S, AF154114); Bt, *Bos taurus* (cow; PGRP-S, AY083309); Cd, *Camelus dromedarius* (camel; PGRP-S, AJ131676); Ss, *Sus scrofa* (pig; PGRP-S, AY598969); Dm, *D. melanogaster* (fruit fly; PGRP-SA, AF207541; PGRP-LB, AF207537). For human PGRP-Iα, C indicates the C-terminal PGRP domain. Sequence alignments were performed using the program ClustalW.[39] The Figure was generated using ESPript.[40] (b) Superposition of the PGRP-specific segments of human PGRP-S (green) and PGRP-IαC (yellow). Residues are labeled as PGRP-S/PGRP-IαC. The segments were superposed through the main-chain atoms of their conserved central regions (residues 15–21 of PGRP-S and 184–190 of PGRP-IαC). (c) Molecular surface of human PGRP-S showing the hydrophobic groove (outlined in yellow) formed by the PGRP-specific segment and helix α2. The protein is oriented similarly as in Figure 1(b). Hydrophobic regions are green; polar regions are red. Surface hydrophobicities were calculated with GRASP.[41] (d) Molecular surface of human PGRP-IαC with its hydrophobic groove formed by PGRP-specific segment outlined in yellow. The orientation is the same as in (a).

mammals appear no more closely related with respect to the topology and dimensions of this putative protein-binding groove than do S and I form PGRPs from the same species. This structural diversity among PGRP-specific segments suggests that individual PGRPs are functionally distinguished by their unique N-terminal regions, which serve to physically link PGN recognition to different signaling pathways or effector molecules.

Protein Data Bank accession code

Coordinates and structure factors for human PGRP-S have been deposited in the RCSB Protein Data Bank under accession code 1YCK.

Acknowledgements

This work was supported by National Institutes of Health Grant AI47900 (to R.A.M.).

References

1. Hoffmann, J. A. (2003). The immune response of *Drosophila*. *Nature*, **426**, 33–38.
2. Dziarski, R. (2004). Peptidoglycan recognition proteins (PGRPs). *Mol. Immunol.* **40**, 877–886.
3. Kang, D., Liu, G., Lundstrom, A., Gelius, E. & Steiner, H. A. (1998). A peptidoglycan recognition protein in innate immunity conserved from insects to humans. *Proc. Natl Acad. Sci. USA*, **95**, 10078–10082.
4. Ochiai, M. & Ashida, M. (1999). A pattern recognition protein for peptidoglycan. Cloning the cDNA and the gene of the silkworm, *Bombyx mori*. *J. Biol. Chem.* **274**, 11854–11858.
5. Werner, T., Liu, G., Kang, D., Ekengren, S., Steiner, H. & Hultmark, D. (2000). A family of peptidoglycan recognition proteins in the fruit fly *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA*, **97**, 13772–13777.
6. Liu, C., Gelius, E., Liu, G., Steiner, H. & Dziarski, R. (2000). Mammalian peptidoglycan recognition protein binds peptidoglycan with high affinity, is expressed in neutrophils, and inhibits bacterial growth. *J. Biol. Chem.* **275**, 24490–24499.
7. Liu, C., Xu, Z., Gupta, D. & Dziarski, R. (2001). Peptidoglycan recognition proteins: a novel family of four human innate immunity pattern recognition molecules. *J. Biol. Chem.* **276**, 34686–34694.
8. Mellroth, P., Karlsson, J. & Steiner, H. (2003). A scavenger function for a *Drosophila* peptidoglycan recognition protein. *J. Biol. Chem.* **278**, 7059–7064.
9. Wang, Z.-M., Li, X., Cocklin, R. R., Wang, M., Wang, M., Fukase, K. *et al*. (2003). Human peptidoglycan recognition protein-L (PGRP-L) is an *N*-acetylmuramoyl-L-alanine amidase. *J. Biol. Chem.* **278**, 49044–49052.
10. Medzhitov, R. & Janeway, C. A., Jr (2002). Decoding the patterns of self and nonself by the innate immune system. *Science*, **296**, 298–300.
11. van Heijenoort, J. (2001). Formation of glycan chains in the synthesis of bacterial peptidoglycans. *Glycobiology*, **11**, 25R–36R.
12. Doyle, R. J. & Dziarski, R. (2001). The bacterial cell: peptidoglycan. In *Molecular Medical Microbiology* (Sussman, M., ed.), pp. 137–153, Academic Press, London.
13. Inohara, N. & Nunez, G. (2003). NODS: intracellular proteins involved in inflammation and apoptosis. *Nature Rev. Immunol.* **3**, 371–382.
14. Chamaillard, M., Girardin, S. E., Viala, J. & Philpott, D. J. (2003). Nods, Nalps and Naips: intracellular regulators of bacterial-induced inflammation. *Cell. Microbiol.* **5**, 581–592.
15. Cheng, X., Zhang, X., Pflugrath, J. W. & Studier, F. W. (1994). The structure of bacteriophage T7 lysozyme, a zinc amidase and an inhibitor of T7 RNA polymerase. *Proc. Natl Acad. Sci. USA*, **91**, 4034–4038.
16. Kaneko, T., Goldman, W. E., Mellroth, P., Steiner, H., Fukase, K., Kusumoto, S. *et al*. (2004). Monomeric and polymeric Gram-negative peptidoglycan but not purified LPS stimulate the *Drosophila* IMD pathway. *Immunity*, **20**, 637–649.
17. Christophides, G. K., Zdobnov, E., Barillas-Mury, C., Birney, E., Blandin, S., Blass, C. *et al*. (2002). Immunity-related genes and gene families in *Anopheles gambiae*. *Science*, **298**, 159–165.
18. Michel, T., Reichhart, J. M., Hoffmann, J. A. & Royet, J. (2001). *Drosophila* Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein. *Nature*, **414**, 756–759.
19. Choe, K. M., Werner, T., Stoven, S., Hultmark, D. & Anderson, K. V. (2002). Requirement for a peptidoglycan recognition protein (PGRP) in relish activation and antibacterial immune responses in *Drosophila*. *Science*, **296**, 359–362.
20. Gottar, M., Gobert, V., Michel, T., Belvin, M., Duyk, G., Hoffmann, J. A. *et al*. (2002). The *Drosophila* immune response against Gram-negative bacteria is mediated by a peptidoglycan recognition protein. *Nature*, **416**, 640–644.
21. Leulier, F., Parquet, C., Pili-Floury, S., Ryu, J.-H., Caroff, M., Lee, W.-J. *et al*. (2003). The *Drosophila* immune system detects bacteria through specific peptidoglycan recognition. *Nature Immunol.* **4**, 478–484.
22. Werner, T., Borge-Renberg, K., Mellroth, P., Steiner, H. & Hultmark, D. (2003). Functional diversity of the *Drosophila* PGRP-LC gene cluster in the response to lipopolysaccharide and peptidoglycan. *J. Biol. Chem.* **278**, 26319–26322.
23. Takehana, A., Katsuyama, T., Yano, T., Oshima, Y., Takada, H., Aigaki, T. & Kurata, S. (2002). Overexpression of a pattern-recognition receptor, peptidoglycan-recognition protein-LE, activates imd/relish-mediated antibacterial defense and the prophenoloxidase cascade in *Drosophila* larvae. *Proc. Natl Acad. Sci. USA*, **99**, 13705–13710.
24. Takehana, A., Yano, T., Mita, S., Kotani, A., Oshima, Y. & Kurata, S. (2004). Peptidoglycan recognition protein (PGRP)-LE and PGRP-LC act synergistically in *Drosophila* immunity. *EMBO J.* **23**, 4690–4700.
25. Kim, M.-S., Byun, M. & Oh, B.-H. (2003). Crystal structure of peptidogylcan recognition protein LB from *Drosophila melanogaster*. *Nature Immunol.* **4**, 787–793.
26. Gelius, E., Person, C., Karlsson, J. & Steiner, H. (2003). A mammalian peptidoglycan protein with *N*-acetylmuramoyl-L-alanine amidase activity. *Biochem. Biophys. Res. Commun.* **306**, 988–994.
27. Dziarski, R., Platt, K. A., Gelius, E., Steiner, H. & Gupta, D. (2003). Defect in neutrophil killing and increased susceptibility to infection with non-pathogenic gram-positive bacteria in peptidoglycan recognition protein-S (PGRP-S)-deficient mice. *Blood*, **102**, 689–697.
28. Tydell, C. C., Yount, N., Tran, D., Yuan, J. & Selsted, M. E. (2002). Isolation, characterization, and antimicrobial properties of bovine oligosaccharide-binding protein. *J. Biol. Chem.* **277**, 19658–19664.
29. Sashchenko, L. P., Dukhanina, E. A., Yashin, D. V., Shatalov, Y. V., Romanova, E. A., Korobko, E. V. *et al*. (2004). Peptidoglycan recognition protein Tag7 forms

a cytotoxic complex with heat shock protein 70 in solution and in lymphocytes. *J. Biol. Chem.* **279**, 2117–2124.
30. Reiser, J.-B., Teyton, L. & Wilson, I. A. (2004). Crystal structure of the *Drosophila* peptidoglycan recognition protein (PGRP)-SA at 1.56 Å resolution. *J. Mol. Biol.* **340**, 909–917.
31. Chang, C. I., Pili-Floury, S., Herve, M., Parquet, C., Chelliah, Y., Lemaitre, B. *et al*. (2004). A *Drosophila* pattern recognition receptor contains a peptidoglycan docking groove and unusual L,D-carboxypeptidase activity. *PLoS Biol.* **2**, 1293–1302.
32. Guan, R., Malchiodi, E. L., Wang, Q., Schuck, P. & Mariuzza, R. A. (2004). Crystal structure of the C-terminal peptidoglycan-binding domain of human peptidoglycan recognition protein Iα. *J. Biol. Chem.* **279**, 31873–31882.
33. Guan, R., Roychowdhury, A., Ember, B., Kumar, S., Boons, G.-J. & Mariuzza, R. A. (2004). Structural basis for peptidoglycan binding by peptidoglycan recognition proteins. *Proc. Natl Acad. Sci. USA*, **101**, 17168–17173.
34. Gobert, V., Gottar, M., Matskevich, A. A., Rutschmann, S., Royet, J., Belvin, M. *et al*. (2003). Dual activation of the *Drosophila* Toll pathway by two pattern recognition receptors. *Science*, **302**, 2126–2130.
35. Pili-Floury, S., Leulier, F., Takahashi, K., Saigo, K., Samain, E., Ueda, R. & Lemaitre, B. (2004). *In vivo* RNA interference analysis reveals an unexpected role for GNBP1 in the defense against Gram-positive bacterial infection in *Drosophila* adults. *J. Biol. Chem.* **279**, 12848–12853.
36. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
37. Merrit, E. A. & Bacon, D. J. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524.
38. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al*. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
39. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.
40. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). ESPript: multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305–308.
41. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.
42. Pflugrath, J. W. (1999). The finer things in X-ray diffraction data collection. *Acta Crystallog. sect. D*, **55**, 1718–1725.
43. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A*, **50**, 157–163.
44. McRee, D. E. (1999). XtalView/Xfit. A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165.
45. Lamzin, V. S. & Wilson, K. S. (1993). Automated refinement of protein models. *Acta Crystallog. sect. D*, **49**, 129–147.
46. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–253.

Edited by I. Wilson

(Received 23 December 2004; received in revised form 24 January 2005; accepted 26 January 2005)