doi:10.1016/j.jmb.2008.03.018 *J. Mol. Biol.* (2008) **378**, 923–932

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Crystal Structure of the Peptidoglycan Recognition Protein at 1.8 Å Resolution Reveals Dual Strategy to Combat Infection Through Two Independent Functional Homodimers

Pradeep Sharma[1], Nagendra Singh[1], Mau Sinha[1], Sujata Sharma[1], M. Perbandt[2], C. Betzel[2], Punit Kaur[1], A. Srinivasan[1] and Tej P. Singh[1]*

[1]*Department of Biophysics All India Institute of Medical Sciences, New Delhi, India*

[2]*Department of Biochemistry and Molecular Biology University of Hamburg Hamburg, Germany*

Received 28 January 2008; received in revised form 5 March 2008; accepted 11 March 2008
Available online 19 March 2008

The mammalian peptidoglycan recognition protein-S (PGRP-S) binds to peptidoglycans (PGNs), which are essential components of the cell wall of bacteria. The protein was isolated from the samples of milk obtained from camels with mastitis and purified to homogeneity and crystallized. The crystals belong to orthorhombic space group *I*222 with $a=87.0$ Å, $b=101.7$ Å and $c=162.3$ Å having four crystallographically independent molecules in the asymmetric unit. The structure has been determined using X-ray crystallographic data and refined to 1.8 Å resolution. Overall, the structures of all the four crystallographically independent molecules are identical. The folding of PGRP-S consists of a central β-sheet with five β-strands, four parallel and one antiparallel, and three α-helices. This protein fold provides two functional sites. The first of these is the PGN-binding site, located on the groove that opens on the surface in the direction opposite to the location of the N terminus. The second site is implicated to be involved in the binding of non-PGN molecules, it also includes putative N-terminal segment residues (1–31) and helix α2 in the extended binding. The structure reveals a novel arrangement of PGRP-S molecules in which two pairs of molecules associate to form two independent dimers. The first dimer is formed by two molecules with N-terminal segments at the interface in which non-PGN binding sites are buried completely, whereas the PGN-binding sites of two participating molecules are fully exposed at the opposite ends of the dimer. In the second dimer, PGN-binding sites are buried at the interface while non-PGN binding sites are fully exposed at the opposite ends of the dimer. This form of dimeric arrangement is unique and seems to be aimed at enhancing the capability of the protein against specific invading bacteria. This mode of functional dimerization enhances efficiency and specificity, and is observed for the first time in the family of PGRP molecules.

Keywords: crystal structure; antimicrobial activity; pattern recognition protein; innate immunity; peptidoglycan recognition protein

Edited by R. Huber

**Corresponding author.* E-mail address: tps_aiims@hotmail.com.

Abbreviations used: PGRP, peptidoglycan recognition protein; PGN, peptidoglycan; CPGRP, camel peptidoglycan recognition protein; PGRP-S, short peptidoglycan recognition protein; PGRP-L, long peptidoglycan recognition protein; PGRP-I, intermediate peptidoglycan recognition protein; HSP, heat shock protein; HPGRP, human peptidoglycan recognition protein; DPGRP, Drosophila peptidoglycan recognition protein; DAP, di-aminopimelic acid.

Introduction

Peptidoglycan recognition proteins (PGRPs) are pattern recognition molecules of the innate immune system. They bind to bacterial peptidoglycan (PGN) molecules with high affinities and terminate the resulting cell–cell communication to protect the host against bacterial infections.[1] PGNs are an integral part of the cell walls of almost all known bacteria. PGNs are polymers that consist of *N*-acetylglucosamine (NAG) and *N*-acetylmuramic acid (NAM) in β (1–4) linkages with short peptide stems consisting of alternating L- and D-amino acids.[2] PGRPs are also structurally related to bacteriophage T7 lysozyme[3] and are broadly classified into three main groups: short PGRP (PGRP-S) with molecular mass ranging from 20–25 kDa; intermediate PGRP (PGRP-Iα and Iβ) with molecular mass of 40–45 kDa; and the long PGRP (PGRP-L) with molecular mass up to 90 kDa.

PGRP-S is a soluble, conserved pattern recognition protein that binds to bacterial PGN with a high affinity. It is a bactericidal protein with specificities for both Gram-negative and Gram-positive bacteria. In order to carry out these functions, PGRP-S has evolved novel structural features, including a highly conserved PGN recognition binding site and a variable non-PGN binding site.[4] Our understanding of the mode of binding of PGRPs depends largely on the knowledge of their three-dimensional structures. Such knowledge is still limited to a few cases. So far, crystal structures of PGRP-LB,[5] PGRP-SA,[6] PGRP-LC[7] and PGRP-LE[8] from *Drosophila* and C-terminal PGN-binding domain of PGRP-IαC[9] and a cloned truncated PGRP-S[4] from human has been reported. We report here the first crystal structure of the secretory PGRP-S at 1.8 Å resolution. It was isolated from mammary secretions of camels with mastitis. The concentration of this protein is reported to be highest in milk during mastitis infection, and it is reported to have a concentration of about 120 mg/l during early lactation.[10] However, thereafter the level declines slowly to ~19% during late lactation but it increases rapidly by 45% during the mastitis infection. This protein is known to bind to lactic acid bacteria and other Gram-positive bacteria with an affinity comparable to that reported for human and murine PGRPs.[10] One of the most characteristic features of mammalian PGRP-S is its high isoelectric point of 9.02[10] as compared to that of insect PGRP-S, which has an acidic pI value of 6.2. The structure has revealed that the protein exists in the form of two independent dimers, each of which has two fully accessible binding sites, one for binding to PGNs and one for the binding to non-PGN molecules.

Results

Quality of the model

The full polypeptide chain of PGRP-S from camel (CPGRP-S) contains residues 1–171. The final structural model is composed of four such CPGRP-S molecules in the crystallographic asymmetric unit. This corresponds to the model having 5348 independent protein atoms. In addition, the final model contains 13 glycerol molecules, one tartrate molecule, one sulphate ion and 739 water oxygen atoms. The overall mean *B*-factor for all -atoms is 36.6 Å^2, which is higher than the values reported for other PGRP structures.[4–6,9] However, the polypeptide chains of all the four molecules are well defined in the electron densities. A Ramachandran plot[11] calculated for the final model with program PROCHECK[12] shows that 87.9% of the residues are in the most favoured regions of the ϕ,ψ plot, with the remaining residues all falling into the additionally allowed regions. The final refinement statistics are given in Table 1.

Table 1. Data collection and refinement statistics

PDB ID	3C2X
Space group	*I*222
Unit cell dimensions	
a (Å)	87.0
b (Å)	101.7
c (Å)	162.3
No. molecules in the unit cell	32
Resolution range (Å)	20.0–1.8
No. measured reflections	39,5870
No. unique reflections	65,264
*R*sym (%)	6.5 (26.1)
$I/\sigma(I)$	14.2 (5.1)
Overall completeness of data (20.0–1.8 Å(%)	99.8 (98.8)
R_{cryst} (%)	22.5
R_{free} (%) 2.5% of reflections	24.7
Protein atoms	5348
Water oxygen atoms	739
Atoms of glycerol (13 molecules)	78
Atoms of tartrate (one molecule)	10
Atoms of sulphate ion (one molecule)	5
r.m.s.d from ideal	
Bond lengths (Å)	0.01
Bond angles (°)	2.0
Torsion angles (°)	24.5
Wilson's *B*-factor (Å^2)	31.0
Mean *B*-factor	
Main chain atoms (Å^2)	34.4
Side chain and water atoms (Å^2)	38.3
All atoms(Å^2)	36.6
Ramachandran ϕ,ψ map	
Residues in the most favoured regions (%)	87.9
Residues in the additionally allowed regions (%)	12.1

The numbers in parentheses correspond to the data in the highest resolution shell.

Overall molecular structure

The four crystallographically independent molecules in the structure are designated as A, B, C and D (Fig. 1). These are associated in the form of two independent dimers; dimer 1 and dimer 2. Dimer 1 contains molecules A and B arranged in a back-to-back orientation with an approximate 2-fold rotation axis (Fig. 1a) while dimer 2 is composed of molecules C and D with face-to-face orientation in which molecules C and D are also related by 2-fold