The Crystal Structure of the Bacteriophage PSA Endolysin Reveals a Unique Fold Responsible for Specific Recognition of Listeria Cell Walls

Ingo P. Korndörfer1⁎, Joseph Danzer1, Mathias Schmelcher2, Markus Zimmer2, Arne Skerra1⁎ and Martin J. Loessner2

1Technische Universität München, Lehrstuhl für Biologische Chemie An der Saatzucht 5 D-85350 Freising Germany
2Institute of Food Science and Nutrition, ETH Zurich Schmelzbergstrasse 7 CH-8092 Zürich Switzerland

Bacteriophage murein hydrolases exhibit high specificity towards the cell walls of their host bacteria. This specificity is mostly provided by a structurally well defined cell wall-binding domain that attaches the enzyme to its solid substrate. To gain deeper insight into this mechanism we have crystallized the complete 314 amino acid endolysin from the temperate Listeria monocytogenes phage PSA. The crystal structure of PlyPSA was determined by single wavelength anomalous dispersion methods and refined to 1.8 Å resolution. The two functional domains of the polypeptide, providing cell wall-binding and enzymatic activities, can be clearly distinguished and are connected via a linker segment of six amino acid residues. The core of the N-acetylmuramoyl-L-alanine amidase moiety is formed by a twisted, six-stranded β-sheet flanked by six helices. Although the catalytic domain is unique among the known Listeria phage endolysins, its structure is highly similar to known phosphorylase/hydrolase-like α/β-proteins, including an autolysin amidase from Paenibacillus polymyxa. In contrast, the C-terminal domain of PlyPSA features a novel fold, comprising two copies of a β-barrel-like motif, which are held together by means of swapped β-strands. The architecture of the enzyme with its two separate domains explains its unique substrate recognition properties and also provides insight into the lytic mechanisms of related Listeria phage endolysins, a class of enzymes that bear biotechnological potential.

© 2006 Elsevier Ltd. All rights reserved.

*Corresponding authors

Keywords: phosphorylase/hydrolase-like α/β-proteins; zinc protease; peptidoglycan; gene duplication; domain swapping

Introduction

Listeria monocytogenes is an opportunistic pathogen responsible for severe infections in both animals and humans. Listeriosis usually manifests itself by symptoms such as meningitis, eye or skin infections, but can also lead to miscarriages.1,2 Human victims are mostly pregnant women, neonates, and immuno-compromised adults. It is now clear that outbreaks of listeriosis are generally food-borne, due to consumption of contaminated meat, dairy products or vegetables.3,4

Bacteriophage-encoded peptidoglycan hydrolases are cell wall lytic enzymes that are synthesized late during virus replication, mediating release of progeny virions. These enzymes were thus termed endolysins, and they are generally able to attack glycosidic linkages or the peptide bridges that cross-
link the glycan strands in the bacterial murein sacculus. These Listeria phage endolysins were first investigated more than ten years ago and have found a number of biotechnological applications since then. These enzymes show not only extraordinary substrate specificity but also high activity when added externally to the Gram-positive Listeria cells, leading to rapid degradation of the murein layers and lysis of susceptible cells.

Based on their antimicrobial properties, endolysins from phages infecting Gram-positive pathogens have recently attracted attention as potential therapeutic agents, particularly in topical applications. Initial studies were successfully carried out with oral Streptococci in mice as well as with Bacillus anthracis in vitro. Endolysins may further be used for surface decontamination purposes or for biopreservation of food and feed. As an example,

**Figure 1.** Alignment of the PlyPSA amino acid sequence with the sequences of Ply500 and of the catalytic domain of CwV (PDB accession code 1JWQ). The DSSP secondary structure assignment for the PlyPSA crystal structure is shown above its sequence. Residue numbering corresponds to that of native PlyPSA (without the Strep-tag II, which is depicted in lower case letters), as it was also used for the structural model. Residues identical to PlyPSA are boxed in grey. p denotes the sequence of the proximal subdomain of the PlyPSA CBD (with respect to the EAD, c.f. Figure 8), which has been structurally aligned with the distal domain, d. The sequences of the distal and proximal subdomains are coloured in red and orange, respectively, with the exception of the structurally swapped β-strands. These are coloured according to their location on the plyPSA gene, i.e. indicating their likely position in the ancestral structure “before” swapping, and marked by boxes.
genetically modified lactic acid bacteria that are able to synthesize and secrete active Listeria phage endolysin were constructed to protect food fermentation products. Hence, endolysins may be regarded as natural, highly specific antimicrobials, and they show a number of properties that are advantageous for corresponding applications: (i) peptidoglycan and carbohydrate scaffolds of the cell walls from Gram-positive bacteria cannot easily be modified via mutagenesis of the bacteria, making development of resistance unlikely; (ii) the high biological specificity permits specific elimination of the target pathogen while the natural bacterial flora or the recombinant host cells themselves are not affected; (iii) the enzymes generally show a modular design comprising separate enzymatically active (EAD) and cell wall-binding domains (CBD). The CBDs serve to specifically recognize and bind the bacterial surface, which defines target specificity largely independent of the catalytic mechanism.

In fact, different EADs and CBDs can be combined as building blocks in order to create hybrid fusion proteins and to optimize the lytic effects on selected target cells (M. S. et al., unpublished results). A similar combination of functional modules has been demonstrated for other endolysins and cell wall hydrolases, e.g. from Streptococcus pneumoniae phages, Staphylococcus aureus phages or Bacillus cereus and B. anthracis prophages. Although both the EAD and CBD seem to be required for lytic activity, isolated C-terminal domains retain their ability to bind to the bacterial cell walls with nanomolar affinity.

So far, little is known about the corresponding ligand structures on the bacterial cell walls. At least in Gram-positive bacteria these conserved motifs appear to comprise mostly carbohydrates. With respect to the Listeria phage CBDs, the unique structure of the teichoic acid polyribitol-phosphate backbone may be important. However, inspection of their amino acid sequences did not reveal any known cell surface anchoring signature.

As a first step towards understanding the molecular basis for cell wall recognition by Listeria phage endolysin, we have solved the crystal structure of PlyPSA, an N-acetyl-muramyl-L-alanine encoded by the phage PSA, and functionally characterized its two domains.

Results

Crystallization, structure determination, and quality of the final model

The full length 314 amino acid PlyPSA endolysin was produced in the cytosol of Escherichia coli as a fusion protein with an N-terminal Strep-tag II affinity peptide (Figure 1). The recombinant protein was purified from the whole cell extract via streptavidin affinity chromatography and gel filtration, with a final yield of approximately 5 mg per 1 l bacterial culture. Crystals of PlyPSA (with the Strep-tag II still attached) were grown at pH 7.8 in the presence of polyethylene glycol (PEG) 3350 as the main precipitant. The crystal structure was solved by single wavelength anomalous dispersion (SAD) phasing using crystals of the seleno-methionine (SeMet) derivative. The initial model was refined against a native data set to a resolution of 1.8 Å with a final R-factor of 0.177 and good overall stereochemistry (Table 1).

Eleven residues at the N terminus of the protein, comprising the Strep-tag II, are disordered and thus not included in the structural model. The active site is most likely occupied by a Zn^{2+}, which is coordinated by N\textsuperscript{ε1} of His10 (2.0 Å), N\textsuperscript{ε1} of His79 (2.1 Å), and O\textsuperscript{δ1} and O\textsuperscript{δ2} of Glu23 (2.8 and 2.1 Å, respectively). Additional electron density in the active site was modelled as a putative peptide ligand of two residues (Glu and Lys). Main chain O and N of the Glu residue from this dipeptide contribute to the approximately trigonal bipyramidal coordination sphere of the Zn\textsuperscript{2+} (2.3 Å and 2.1 Å, respectively (Figure 2), in accordance with previously described coordination geometries for Zn\textsuperscript{2+}. In the current model, average B-factors are 34.6 Å\textsuperscript{2} for the protein chain and 39.9 Å\textsuperscript{2} for the presumed dipeptide ligand, while average B-factors for solvent molecules and bound ions (see Materials and Methods) are 47.0 Å\textsuperscript{2}.

Overall structure and functional domains of PlyPSA

PlyPSA is composed of two structurally distinct domains that are connected by a short linker peptide of six amino acid residues (Figure 3). The N-terminal domain comprises residues 1–176 and was earlier proposed to carry the enzymatic active site of the endolysin, based on sequence homology with other enzymes of the amidase type. The C-terminal domain, comprising residues 183–314, was therefore assumed to harbour the cell wall-binding function (Figure 1).

To confirm the location of the catalytic function, a PlyPSA fragment (Met1 to Ser185) encompassing the entire N-terminal EAD but lacking the presumed CBD was cloned and produced in E. coli, purified, and finally tested for lytic activity in a photometric

---

**Table 1. Refinement statistics**

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>2.50–1.80 (1.85–1.80)*</td>
</tr>
<tr>
<td>Reflections used</td>
<td>46,406 (3388)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (89.4)*</td>
</tr>
<tr>
<td>R/WRfree</td>
<td>0.170/0.198 (0.202/0.228)*</td>
</tr>
<tr>
<td>Protein residues visible</td>
<td>315</td>
</tr>
<tr>
<td>No. of solvent molecules</td>
<td>307</td>
</tr>
<tr>
<td>r.m.s.d. from ideality</td>
<td>0.014</td>
</tr>
<tr>
<td>of bond lengths (Å)</td>
<td>1.591</td>
</tr>
<tr>
<td>r.m.s.d. from ideality of bond angles (°)</td>
<td>1.91</td>
</tr>
</tbody>
</table>

* Numbers in parenthesis are for the outermost resolution shell.
lysis assay based on the decrease in turbidity of a bacterial substrate suspension (Figure 4). The isolated EAD revealed a detectable lytic activity of 81 units/mg, thus demonstrating its catalytic function. However, under the same experimental conditions the complete PlyPSA enzyme exhibited an approximately sevenfold higher activity (543 units/mg), which indicates that the C-terminal CBD is required for full activity of the enzyme on its natural substrate.

In a separate experiment the C-terminal CBD was shown to be responsible for specific targeting to the Listeria cell wall. A fusion between the green fluorescent protein (GFP) and the PlyPSA fragment comprising residues Thr172 to Lys314 was prepared and tested for cell wall decoration of a comprehensive set of Listeria strains as described before.\textsuperscript{13} The fusion protein was shown to uniformly decorate the cell walls of serovars 4 (L. Monocytogenes, Figure 4), 5 (L. ivanovii), and 6 (L. innocua), whereas it did not bind to serovars 1/2 or 3. This result corresponds to both the known specificity of the full-length bacteriophage PlyPSA endolysin for the respective L. monocytogenes strains and to the binding pattern observed for the related CBD of the Ply500 endolysin (Ply500\textsuperscript{13}).

**Structure of the enzymatically active domain**

The core of the EAD appears as a heavily twisted, six-stranded β-sheet displaying a ‘+1x, −2x, −2, −1x, +2’ topology\textsuperscript{23} (Figure 3), which is covered from both sides by six α-helices. A search with DALI\textsuperscript{24} identified structural relationship with the catalytic domain of CwIV (PDB accession code 1JWQ; Z-score=22.4), a cell wall lytic N-acetylmuramoyl-l-alanine amidase from Paenibacillus polymyxa var. colistinus. Although the catalytic domain of CwIV displays a sequence identity of merely 21.0% (37 from 176 residues) to PlyPSA the two amidases can be superpositioned with an overall r.m.s.d. of 1.61 Å (148 C\textsuperscript{α}-atoms, Figure 5). The r.m.s.d. of the central β-sheets of the two proteins is even lower at 0.89 Å (44 C\textsuperscript{α}-atoms).

Biochemical studies and site-directed mutagenesis of 20 residues in CwIV that are highly conserved among the CwIB amidase family indicated that CwIV is a zinc-dependent peptidase.\textsuperscript{25} In CwIV the catalytically active Zn\textsuperscript{2+} is coordinated by residues His10, Glu26, His80, and Glu142. These amino acids are located near the C terminus of the three central β-strands (β1, β3, and β6) and are thus presented on a concave platform formed by the β-sheet and the loops emerging from those strands (Figure 5). The corresponding residues in PlyPSA, His10, Glu23, His79, and Glu141, are structurally conserved, and a bound metal ion was clearly identified in the electron density (see Materials and Methods), indicating that PlyPSA constitutes a zinc peptidase, too.

The active site of PlyPSA is formed by an elongated groove on the molecular surface of the EAD (Figure 6(a)). The catalytically active Zn\textsuperscript{2+} assumes a position not quite at the bottom but rather at one of the walls of this depression. Adjacent to this site the EAD possesses a second and considerably deeper pocket (approximately 13 Å×10 Å×8 Å). While the wall of this pocket that is closer to the metal binding centre is lined by both main chain atoms and polar side-chains (e.g. the Zn\textsuperscript{2+}-binding His79), the opposing wall displays predominantly hydrophobic side-chains (e.g. Leu60, Ile63, Phe129, Leu139). Ser7 and Ser77 form the bottom of this pocket.

A pyrrolidone carboxyl peptidase from Thermococcus litoralis (PDB accession code 1A2Z\textsuperscript{26}; Z-score from DALI=8.6) as well as duck carboxypeptidase D domain II (CPD-2; PDB accession code 1QMU\textsuperscript{27}; Z-score from DALI 8.5, Figure 6(b) and (c)) were identified as further proteins with high structural similarity to the EAD of PlyPSA. However, the pyrrolidone carboxyl peptidase is a cysteine protease with a structurally very different active site. Moreover, the region corresponding to the aforementioned, second pocket in PlyPSA is completely covered by a large α-helix. In contrast, CPD-2 is a zinc peptidase similar to PlyPSA and CwIV. It exhibits an N-terminal α/β-hydrolase subdomain with overall topological similarity and general...
coincidence of the key catalytic residues (His74, Glu77, Arg135, Asn144, Arg145, His181, Tyr250, Glu272) with those described for the archetypal pancreatic carboxypeptidase A. The central β-sheets of the catalytically active domains of CPD-2 and PlyPSA can be superpositioned with an r.m.s.d. of 1.19 Å (41 Cα-atoms, Figure 6(b)). CPD-2 and PlyPSA coordinate the Zn\(^{2+}\) in a very similar manner (Figure 6(d)), using a set of homologous side-chains (i.e. His10, Glu23, His79, Glu141 in PlyPSA and His74, Glu77, His181, Glu272 in CPD-2). In CPD-2, Arg135 has been suggested to polarize the scissile substrate bond, and the formation of the oxyanion hole may be assisted by Asp142, which forms a salt-bridge with the guanidinium group of Arg135.

However, in the crystal structure of PlyPSA no structural counterparts for these two residues could be identified. The long loop (residues 113–160) that carries residues Arg135 and Asp142 of CPD-2 is replaced by a very short linker (residues 48–56) in PlyPSA, directly connecting strand β2 with helix α2 and opening access to the deep pocket adjacent on the active site described above. The space occupied by Arg135 of CPD-2 seems to be adopted by the presumed peptide ligand in the PlyPSA structure. Finally, CPD-2 possesses a funnel-like entrance to the active site with a deep specificity pocket close to

Figure 3. Crystal structure of PlyPSA. The EAD is shown in blue, the linker region in grey, and the CBD in red. The catalytic Zn\(^{2+}\) is depicted as a yellow sphere.

Figure 4. Functional characterization of PlyPSA domains. (a) Cell wall lytic activity of native PlyPSA in comparison with the C-terminally truncated enzyme (i.e. the isolated EAD, residues 1–185). A suspension of L. monocytogenes cells was used as substrate and the decrease in absorbance at 600 nm (A\(_{600}\)) was monitored after addition of equimolar amounts of the proteins (or buffer, control). Values were normalized to an A\(_{600}\) of 1.0. (b) Targeting of GFP-tagged CBD to Listeria cell walls. The GFP-CBD fusion protein recognizes and binds to an evenly distributed ligand on the surface of L. monocytogenes serovar 4 cells, as visualized by epifluorescence microscopy.
the catalytic centre (Figure 6(c)). The long loop formed by residues 246 to 257, which defines a large part of the S1’ sub-site in CPD-2 and accepts the C-terminal residue of its peptide substrate, is shorter in PlyPSA (residues 123 to 130) again. Thus, the entrance to the active site in the EAD of PlyPSA should be generally better accessible by the peptidoglycan substrate of this endolysin.

The bound putative substrate ligand

While the biochemical identity of the bound dipeptide ligand visible in the electron density of the crystallized PlyPSA could not be experimentally confirmed, the amino group and the main chain oxygen of the N-terminal residue, possibly glutamate, interact closely with the catalytic Zn$^{2+}$. With increasing distance from the metal ion, the extra density observed for the probably longer peptide ligand (cf. Figure 6(a)) becomes less pronounced such that only two residues could be modelled with confidence in the form of a representative dipeptide.

Although it is surprising that, following several dialysis and gel filtration steps, a low molecular mass ligand was still present in the crystallized protein preparation, the bound peptide fragment could either constitute a muropeptide liberated from E.coli during cell lysis or it might originate from contamination with host cell proteases or even from partial autoproteolytic degradation of PlyPSA. In fact, we observed that the purified endolysin had to “incubate” for four weeks at 4 °C before crystallization was possible. After this time, truncated protein fragments were detectable via SDS PAGE (data not shown), potentially resulting in peptide fragments able to bind to the active site. Indeed, removal of the proteolytic polypeptide fragments from these samples by gel filtration seemed to prevent the formation of crystals.

Structure of the linker and interface between the two functional domains

EAD and CBD of PlyPSA are connected via a short linker fragment (residues 177–182; Figures 3 and 7). While the linker fragment forms one hydrogen bond each to both the EAD (Ala177/N → Glu41/Oε2) and to the CBD (Asn182/Nδ2 → Asn291/Oε1), only one direct polar interaction was observed between the EAD and CBD (Trp305/Nε1 → Arg40/O). Apart from this, only the aromatic side-chains of residues Trp299 and Trp305 of the CBD are in contact with the carbon backbone of Arg40 in the EAD.

Structure of the cell wall-binding domain

The CBD of PlyPSA exhibits a novel fold. Despite the lack of any recognizable pattern of internal amino acid sequence similarities, such as duplicated motifs or repeats, the CBD is obviously composed of two structurally homologous subdomains (Figure 8(a)). Closer inspection after mutual superposition of both subdomains revealed that the (sequence-wise) “leading” β-strand of the proximal domain (with respect to the N-terminal EAD) and its “trailing” β-strand of the distal domain, are structurally swapped between the two domains (Figure 8(a) and (b), cf. Figure 1).

Thus, the observed structure could be the result of either a gene duplication during evolution of the CBD or the pick-up of another functionally equivalent coding sequence, followed by swapping of the respective “ancestral” leading β-strands (β1p and β1p). In the ancestral proximal subdomain a first strand-loop-strand motif (β1p-loop-β2p) would be followed by a second strand-loop-strand motif (β3p-loop-β4p) which folds back onto the first one and thereby leads to a pair (i.e. β1p, with β4p, and β2p with β3p) of two-stranded β-sheets. A consecutive β-hairpin (β5p-loop-β6p) would wedge itself between these two sheets in an almost orthogonal manner and keep them apart (Figure 8(c)).

The ancestral domain would be terminated by a short helix (αi5) and a final short strand (βi5) motif. This architecture is essentially maintained after swapping of the aforementioned β-strands, which is indicated by the fact that the two subdomains of the PlyPSA CBD can be superpositioned with an r.m.s. s.d. of 1.73 Å for 56 Cα-atoms (Figure 8(b); residues 190–228, 231–241, and 251–256 from the distal domain with residues 183–188, 262–300, and 304–314 from the proximal domain). Sequence identity between the two domains, even after structural alignment (LSQMAN96), is surprisingly low at just 19%.

Notably, the CBD features a high content of aromatic amino acids (14 Tyr, three Phe, and six...
Trp residues of 135 residues in total). Side-chains of these residues contribute an exposed surface area of 780 Å² as opposed to just 297 Å² for residues of the same kind (four Tyr, three Phe, zero Trp residues of 176 residues) in the EAD (Figure 8(d)). A distinctive patch of these hydrophobic residues forms part of a cleft at the interface between the two structural subdomains of the CBD (residues Phe200, Tyr222, Trp223, Tyr236, Phe239, and Trp279; Figure 8(d)), which might indicate a binding region for an oligosaccharide ligand. The suitability of this region for ligand binding is confirmed by a prediction of energetically favourable binding sites with Q-SiteFinder (data not shown).

**Discussion**

Here we describe the first three-dimensional structure of a murein hydrolase from a *Listeria*...
bacteriophage. Structural and functional characterization of the PlyPSA endolysin confirms the modular assembly of this class of enzymes, which was previously predicted on the basis of sequence alignments with various peptidoglycan hydrolases of prokaryotic origin.\textsuperscript{13,18} This type of modular design allows for an independent biological evolution of the two domains and facilitates exchange of either of the two domains in response to a selective pressure. Furthermore, it permits the design and construction of artificial hybrid enzymes with novel properties and tailored for biotechnological applications (M. S. et al., unpublished results).

The structural topology of the EAD led to its classification as a member of the phosphorylase/hydrolase-like $\alpha/\beta$-proteins. The N-terminal N-acetylmuramoyl-L-alanine amidase moiety of PlyPSA shares up to 40\% amino acid sequence identity with the amidase endolysins from \textit{B. cereus} phage 12826 and \textit{Clostridium perfringens} phage $\phi$3626,\textsuperscript{18,31} amidases from members of the \textit{Bacillus cereus} group, \textit{Clostridium} spp., and from other bacteria.\textsuperscript{18} The pronounced structural homology observed here between the catalytic region of PlyPSA and other known enzymes from this class, in particular CwlV$^{25}$ and CPD-2,\textsuperscript{27} also support the finding that PlyPSA is a zinc peptidase. Indeed, removal of divalent cations with EDTA clearly diminishes the enzymatic activity of PlyPSA (results not shown) even though Zn(II) might be replaced by another suitable divalent metal ion \textit{in vivo}.

We have found that molecular details of the active site differ in some aspects from previously described proteins of this class (i.e. CPD-2). In fact, the catalytic mechanism of PlyPSA is not yet fully understood. Zinc-dependent peptidases usually possess a Glu residue that acts as general base during the nucleophilic attack of a water molecule on the carbonyl group of the scissile peptide bond.\textsuperscript{32} While this residue is conserved as Glu141 in PlyPSA, positively charged residues that could provide an oxyanion hole are absent. However, it cannot be fully ruled out that binding of the presumed peptide

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure7.png}
\caption{Interface between EAD (blue) and CBD (red) with the linker peptide (grey). Polar contacts between the two domains or with the linker are shown as broken lines with labels indicating distances. Residues involved in direct hydrophobic contacts between the two domains are labelled.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure8.png}
\caption{Architecture and putative ligand-binding sites of the PlyPSA CBD. (a) Organisation of its two subdomains. Proximal and distal subdomain of the CBD are coloured in orange and red, respectively, except for the structurally swapped strands, which are coloured according to their sequential neighbourhood (see also Figure 1). (b) Superposition of the proximal and distal subdomains (coloured as in (a)). (c) Schematic representation of the ancestral CBD subdomain. (d) Distribution of aromatic side-chains in the CBD and indication of a putative binding site. Residues with uncharged aromatic side-chains are shown as green stick models (backbone coloured as in (a)). (e) Surface representation of the CBD with hydrophobic surface areas coloured in green (as described\textsuperscript{53}). The white ellipsoid indicates the location of a cleft between the two subdomains that may be suited to bind cell wall associated ligands.}
\end{figure}
ligand observed in our crystal structure leads to a rearrangement of the surrounding loops, thereby shifting the corresponding side-chains. The modelled orientation of the bound dipeptide does not allow a definite conclusion about how the peptidoglycan peptide bond might be attacked in the hydrolytic reaction mechanism. In our current model, where the N-terminal amino group of the dipeptide coordinates the catalytic Zn\(^{2+}\), no space is left for additional amino acids. Hence, this complex probably does not represent the immediate product of a hydrolytic cleavage reaction. A similar “chelating” interaction of a peptide fragment (glycyl-L-tyrosine) with a catalytic Zn\(^{2+}\) was previously observed in carboxypeptidase A (PDB accession code 3CPA\(^{33}\)) and interpreted as “non-productive”.

The deep second pocket observed adjacent to the catalytic site may be suitable for uptake of a terminal sugar unit from the Listeria peptidoglycan. The amphiphilic character of the two opposing walls of this pocket could reflect the mixed hydrophilic and hydrophobic character of a pyranose moiety.\(^{34}\) This would imply that PlyPSA acts preferably on peptide bonds located in vicinity to a terminal sugar residue of the peptidoglycan scaffold, which could contribute to its substrate specificity. However, experimental evidence for such a hypothetical behaviour is not yet available.

Our present finding that the cell wall lytic activity of the PlyPSA EAD is significantly hampered when devoid of its CBD moiety indicates that the latter portion of the protein is required to efficiently direct the enzyme to its macromolecular peptidoglycan substrate. The low residual activity that was detectable for the isolated EAD suggests that the enzymatic hydrolysis of the bacterial cell wall in the lysis assay may be limited by entropic effects. In the absence of the CBD, a successful collision of the enzyme with the peptidoglycan would result in a single cleavage of a peptide crosslink within the multi-layered murein sacculus. In contrast, the presence of the CBD is thought to result in transient, high affinity attachment to the cell wall and may thus permit cleavage of multiple peptide bonds before dissociation. Also, due to the prolonged presence in a specific location of the bacterial peptidoglycan layer a more punctual lysis pattern might result, rather than stochastic lysis of bonds scattered all over the sacculus resulting from repeated encounter.

In this respect, it might be noted that the linker peptide between the EAD and the CBD of PlyPSA is rather short and, at least as deduced from the crystal structure, does not allow for much flexibility between the two domains. Furthermore, the CBD is attached almost opposite to the active site of the EAD. This leads to a different overall topology and conformation of this multidomain protein compared with the cellobiohydrolase I from Trichoderma reesei, for example, another well studied microbial enzyme with a macromolecular substrate and dual domain architecture.\(^{35}\) Cellobiohydrolase I was shown to rely on a long and mobile linker region for optimal activity.\(^{36}\) This probably reflects the differing three-dimensional composition of its substrate, crystalline cellulose, compared with the probably more loosely packed murein multi-layer of the Listeria peptidoglycan.

The CBD of PlyPSA (residues 183–314) features a novel and unique fold, with no significant amino acid sequence homology to other known proteins. The only exception is the closely related C-terminal CBD module encoded by L. monocytogenes phase A500,\(^{13}\) which likely plays an equivalent role. Although the PlyPSA and Ply500 endolysins exhibit different catalytic modules, with N-acetylmuramoyl-l-alanine amidase and l-alanine-f-glutamate endopeptidase activities, respectively, they share 73% identity over 135 residues in their C-terminal cell wall-binding domains. This supports our previous notion that these CBD modules are specifically evolved to recognize a Listeria-specific surface carbohydrate.\(^{13}\) The identity of this ligand, which exclusively occurs on serovar 4, 5, and 6 cells, is still unknown, but preliminary results indicate that the unique structure and carbohydrate decoration of the cell wall teichoic acids is involved. Surprisingly, the PlyPSA CBD structure reveals an internal duplication of a unique fold. The particular arrangement of this motif can be best explained by a domain-swapping mechanism, with structural conservation independent of amino acid sequence. At the interface of its two subdomains the CBD exhibits a pronounced hydrophobic cleft, which is mainly formed by aromatic side-chains and might constitute the substrate recognition site.

Materials and Methods

Cloning, production and purification of PlyPSA for crystallization

The coding sequence for the PlyPSA endolysin was amplified from purified phage DNA as described\(^{19,20}\) and cloned on the E. coli expression vector pASK-IBA5 (IBA, Göttingen, Germany). The resulting plasmid pASK-IBA5-PlyPSA encodes the full length 314 amino acid residue PlyPSA as a fusion protein with an N-terminal Strep-tag II affinity peptide\(^{37}\) of altogether 12 amino acid residues. Cultures (2 l) of E. coli K-12 strain JM83 harbouring pASK-IBA5-PlyPSA were grown in LB medium\(^{28}\) supplemented with 100 mg/l ampicillin at 37 °C. Gene expression was induced at a cell density of \(\Delta A_{550}=0.8\) by adding 0.2 mg/l anhydrotetracycline.\(^{38}\) After further incubation (with shaking) for 3 h, the bacterial cells were harvested by centrifugation, resuspended in 150 mM NaCl, 1 mM EDTA, 100 mM Tris-HCl (pH 8.0) and lysed in a French Press (AMINCO, Rochester, NY). Streptavidin (1.5 mM) was added to mask any biotinylated host cell proteins and the raw extract was then applied to a chromatography column with immobilized engineered streptavidin\(^{39}\) using the same buffer. The NaCl concentration was then gradually increased to 500 mM to elute contaminating proteins associated with the bound PlyPSA and to prevent aggregation of PlyPSA in the following steps. Finally, the recombinant PlyPSA was competitively eluted by application of 2.5 mM D-desthiobiotin (IBA) in the same buffer. Elution fractions...
were concentrated (Amicon Centriprep 10, Millipore, Schwalbach, Germany), applied to a Superdex-75 gel filtration column (Amersham Pharmacia, Uppsala, Sweden), and eluted in a homogenous peak corresponding to the expected size for the monomeric protein. The yield was approximately 5 mg of purified protein per 1 l of *E. coli* culture at 170 KDa . Uniformity and purity of the protein fraction was checked by SDS-PAGE (not shown).

The SeMet derivative of PlyPSA was prepared by repression of the bacterial methionine biosynthesis and complementation of the growth medium. 39 For this purpose, *E. coli* K-12 strain W3110 harbouring pASK-IBA5-PlyPSA was grown in glucose M9 minimal medium, and l-Lys, l-Thr, l-Phe (50 mg/l each), l-Leu, l-Ile, l-Val (25 mg/l each), as well as l-SeMet (25 mg/l) were added at a bacterial cell density of 0.8. After a 15 min incubation, plasmid-mediated gene expression was induced by adding 0.2 mg/ml anhydrotetracycline. The cells were harvested after further shaking at 37 °C for 6 h, and recombinant PlyPSA was purified as above, with the addition of 1 mM dithiothreitol to all buffers. The incorporation of SeMet into the recombinant protein was verified by MALDI-TOF mass spectrometry using a Q-Tof Ultima API (Micromass UK Limited, Manchester, UK) instrument.

### Functional characterization of PlyPSA and its individual domains

The coding sequences for full length PlyPSA protein and its EAD fragment, comprising residues 1–185, were amplified by PCR and cloned on the expression vector pQE-30 (Qiagen, Hilden, Germany). The proteins were produced in *E. coli*, purified by metal affinity chromatography (IMAC), and tested for lytic activity in a photometric lysis assay based on turbidity measurement with intact cells of *L. monocytogenes* WSLC 1042 as substrate as described. 19,20 Identical amounts (2.8 nmol) of each protein were used in the lysis assay. In order to calculate and compare the relative enzymatic activities of the two proteins, the following procedure was applied: photometric curves (\(A_{550\ nmol}\)) were normalized to a common starting value of 1.0 and corrected by subtraction of the corresponding control values (i.e. curves measured without addition of enzyme). Then, the values were fitted with the following function:

\[
y = y_0 + \frac{a}{1 + e^{-c(x-x_0)}} + \frac{b}{1 + e^{-d(x-x_1)}}
\]

The steepest slope of both curves was determined and used for the calculation of endolysin enzymatic activity. 20

A fusion between GFP-His6 and the C-terminal CBD (residues 172 to 314) of PlyPSA was constructed by inserting the gene fragment into the His-GFP backbone of pHGFP as described. 13 The resulting fusion protein was produced in *E. coli*, purified by IMAC, and used to test cell wall binding via decoration of a comprehensive set of *Listeria* cells from different species and serovars. 13

### Crystallography, data collection, and structure determination

Large hexagonal crystals of the native recombinant PlyPSA (carrying the N-terminal Strep-tag II) as well as of its SeMet derivative were grown via the hanging-drop vapour diffusion technique using 25% (w/v) PEG 3350, 0.2 M (NH4)2SO4, 0.1 M Hepes-NaOH (pH 7.8), as precipitant. 1 μl of the protein solution (10 mg/ml in 500 mM NaCl, 10 mM Tris-HCl (pH 8.0)) was mixed with 1 μl of the precipitant solution on a siliconized glass cover slip, followed by equilibration against 0.5 ml of the precipitant solution. Protein crystals were cryo-protected by adding 0.5 μl of 50% (v/v) glycerol to the crystallization drop, harvested using Nylon loops (Hampton Research, Aliso Viejo, CA), and frozen in a 100 K nitrogen stream (Oxford Cryosystems, Oxford, UK). A single wavelength anomalous dataset was collected from one crystal of the SeMet derivative at BM14 (ESRF, Grenoble, France) to a resolution of 2.6 Å (Table 2). Reflection intensities were processed with MOSFLM and SCALA. 41

The space group was determined as P6_122 with one molecule in the asymmetric unit. A dataset of the native protein was also collected at BM14, yielding a resolution of 1.8 Å (Table 2).

Selenium sites were identified in the anomalous dataset using HKL2MAP; 42 Phase calculation resulted in an electron density with features revealing various α-helical elements. Selenium sites were refined with SHARP 43,44 (phasing power=2.89, R_Cut-off=0.46). After solvent flattening with SOLOMON 45 304 residues (out of 315) were modelled automatically using ARP/wARP. 46 Alternating rounds of manual re-building with QUANTA 47 and crystallographic refinement with REFMAC 41 against the native data set led to the final model with an R-factor of 0.214 (Table 1).

During the process of manual model building pronounced \(F_o-F_c\) difference electron densities indicated the presence of several bound ligands. One such peak was interpreted as a bound Zn2+, possibly as a remnant of the bacterial cell extract. Additional density in close proximity was modelled with an exemplary Glu-Lys dipetide (cf. Figures 2 and 6) using an occupancy of 0.8. Tentative modelling of further ligand residues (cf. Figure 6(a)) based

### Table 2. Data collection statistics

<table>
<thead>
<tr>
<th>Space group</th>
<th>P6_122</th>
<th>Native (0.92 Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset</td>
<td>SeMet (peak, 0.98 Å)</td>
<td>90.65, 90.65, 214.09 *</td>
</tr>
<tr>
<td>No. of reflections</td>
<td>16,836 (3344)</td>
<td>48,970 (6932)</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>0.20–2.6 (2.74–2.60) *</td>
<td>25.00–1.80 (1.90–1.80) *</td>
</tr>
<tr>
<td>Completeness of data (%)</td>
<td>100.0 (100.0) *</td>
<td>99.9 (100.0) *</td>
</tr>
<tr>
<td>1/(\sigma(I))</td>
<td>5.6 (1.6) *</td>
<td>5.3 (1.0) *</td>
</tr>
<tr>
<td>Anomalous completeness (%)</td>
<td>100.0 (100.0) *</td>
<td>18.8 (17.0) *</td>
</tr>
<tr>
<td>Redundancy</td>
<td>19.1 (18.2) *</td>
<td>18.8 (17.0) *</td>
</tr>
<tr>
<td>(R_{mean})</td>
<td>0.093 (0.362) *</td>
<td>0.083 (0.716) *</td>
</tr>
</tbody>
</table>

\* Numbers in parenthesis are for the outermost resolution shell.
on some visible electron density at the C-terminal end resulted in significantly higher B-factors such that these residues were not kept in the final model. Further difference electron densities were interpreted as bound SO\textsubscript{4} and Cl\textsuperscript{-} ions and as one Tris molecule, which is involved in a crystallographic contact.

Water molecules were added where stereochemically plausible and the 2\(F_o-F_c\) and \(F_c-F_o\) difference Fourier maps revealed densities of more than 1.0 and 3.5 \(\sigma\), respectively. All backbone dihedral angles fall into allowed regions of the Ramachandran plot.\textsuperscript{18} The only exception is residue Asp19, which is located in a solvent-exposed loop and forms a salt-bridge with the side-chain of Lys134 from a crystallographic neighbour. The electron density for this residue is well defined.

Secondary structure assignments were made with DSSP.\textsuperscript{49} Electrostatic surfaces were calculated with ABPS.\textsuperscript{50} Figures of molecular models, surfaces, and electron densities were prepared with PyMOL\textsuperscript{51} and ABPS TOOLS (M.G. Lerner, University of Michigan).

**Protein Data Bank accession code**

The coordinates of the molecular model of PlyPSA have been deposited at the Protein Data Bank\textsuperscript{52} under accession code 1XOV.

**Acknowledgements**

We thank I. Theobald for providing the streptavidin affinity column, H. Stalz for mass spectrometry of SeMet variants, and the European Molecular Biology Laboratory Grenoble Outstation, in particular H. Belrhali, for supporting the X-ray measurements at the European Synchrotron Radiation Facility under the "Improving Human Potential Programme" of the European Union. We are grateful to S. Scherer for financial support during the initial phase of this project.

**References**


**Edited by R. Huber**

(Received 11 May 2006; received in revised form 17 August 2006; accepted 23 August 2006)

Available online 30 August 2006