LysM, a widely distributed protein motif for binding to (peptido)glycans
Buist, Girbe; Steen, Anton; Kok, Jan; Kuipers, Oscar P.

Published in:
Molecular Microbiology

DOI:
10.1111/j.1365-2958.2008.06211.x

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2008

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
MicroReview

LysM, a widely distributed protein motif for binding to (peptido)glycans

Girbe Buist,†§ Anton Steen,‡§ Jan Kok and Oscar P. Kuipers*
Department of Molecular Genetics, Groningen Biomolecular Sciences and Biotechnology institute, University of Groningen, Kerklaan 30, 9751 NN Haren, the Netherlands.

Summary

Bacteria retain certain proteins at their cell envelopes by attaching them in a non-covalent manner to peptidoglycan, using specific protein domains, such as the prominent LysM (Lysin Motif) domain. More than 4000 (Pfam PF01476) proteins of both prokaryotes and eukaryotes have been found to contain one or more Lysin Motifs. Notably, this collection contains not only truly secreted proteins, but also (outer-)membrane proteins, lipoproteins or proteins bound to the cell wall in a (non-)covalent manner. The motif typically ranges in length from 44 to 65 amino acid residues and binds to various types of peptidoglycan and chitin, most likely recognizing the N-acetylglucosamine moiety. Most bacterial LysM-containing proteins are peptidoglycan hydrolases with various cleavage specificities. Binding of certain LysM proteins to cells of Gram-positive bacteria has been shown to occur at specific sites, as binding elsewhere is hindered by the presence of other cell wall components such as lipoteichoic acids. Interestingly, LysM domains of certain plant kinases enable the plant to recognize its symbiotic bacteria or sense and induce resistance against fungi. This interaction is triggered by chitin-like compounds that are secreted by the symbiotic bacteria or released from fungi, demonstrating an important sensing function of LysMs.

The Lysin Motif

The Lysin Motif (LysM) was first discovered in the lysozyme of Bacillus phage f29, where it was detected as a C-terminal direct repeat composed of 44 amino acids (aa) separated by 7 aa (Garvey et al., 1986). Similar motifs were subsequently observed in the peptidoglycan (PG) hydrolase of Enterococcus faecalis (Béliveau et al., 1991). This enzyme contains a C-terminal LysM domain with six LysMs for which a PG binding function was postulated. Currently, various proteins containing LysMs have been studied (Table S1). LysMs occur frequently in bacterial lysins, in bacteriophage proteins and in certain proteins of eukaryotes (Pfam PF01476 and Prodom PD407905). They are also present in bacterial PG hydrolases and in peptidases, chitinases, esterases, reductases or nucleotidases. They can act as antigens or are proteins that bind, for example, to albumin, elastin or immunoglobulin (Desvaux et al., 2006). Up to now, LysMs have not been found in archaeal proteins (Fig. 1). Multiple LysMs within one LysM domain are separated by spacing sequences mostly consisting of Ser, Thr and Asp or Pro residues (Buist et al., 1995; Ohnuma et al., 2008), which may form a flexible region between the LysMs. The intervening sequences vary in length, composition and do not share significant homology. However, the intervening sequences between the LysMs of a family of receptor-like kinases in plants contain a conserved CxX motif, for which the function is unknown (Madsen et al., 2003; Radutoiu et al., 2003; Arrighi et al., 2006). LysMs are present in the N-terminal as well as the C-terminal domains of proteins; they are also present in the central part of proteins, possibly connecting two (catalytic) domains (Table S1).

The consensus sequence of all available LysMs (Fig. 2) shows that the motif is well conserved over the first 16 and somewhat less over the last 10 amino acid residues. The central region is poorly conserved except for the Ile/Leu at positions 23 and 30 and the well-conserved Asn at...
Prokaryotic LysMs, in contrast to their eukaryotic counterparts, do not possess possible disulphide bridges. Their domains contain extensive secondary structures and hydrogen bond networks, and consequently, disulphide bridges are non-essential for their structure and functions (Ponting et al., 1999). Interestingly, the isoelectric points (pIs) of the LysM proteins range from 4 to 12, with most having a pI of 5 or 10 (Table S1). In some cases, one organism expresses two paralogous LysM domain-containing proteins with different pIs; for example, the homologous PG hydrolases AcmA and AcmD of Lactococcus lactis contain three C-terminal LysMs with pIs of around 10 and 4 respectively. The exact function of LysM domains with low pI is unknown, but they are observed in proteins of several bacteria and could be a way to adapt to changing pHs in the environment or to display an altered substrate binding specificity.

\[ \text{\textbf{N-acetylglucosamine is the general constituent in binding substrates of LysM proteins}} \]

The best-characterized LysM-containing protein is the N-acetylglicosaminidase AcmA of L. lactis (Buist et al., 1995). AcmA binds in a non-covalent manner to the cell wall and is responsible for cell lysis of producing cells as well as lysis in trans (Fig. 3) (Buist et al., 1997). A fusion of a Malaria parasite surface antigen and the C-terminal LysM domain of AcmA bound specifically to cell walls of AcmA-producing and AcmA-non-producing L. lactis cells and to whole cells of several other Gram-positive bacterial species tested (Steen et al., 2003). Binding was obtained in the range of pH 4–10. A similar fusion with the LysM domain of AcmD only bound at a pH below the pI of that domain (~pH 4) (G. Buist et al., unpublished), suggesting that positive charges play an important role in binding. Binding studies using the AcmA fusion protein and chemically treated L. lactis cells and cell walls identified PG as the component to which the LysM domain of AcmA binds (Steen et al., 2003). Binding to purified PG has also been shown for AtlA of E. faecalis, a PG hydrolase containing six C-terminal LysMs (Eckert et al., 2006). The LysM domain of AcmA has similar affinity for both A-type and B-type PG (Steen et al., 2003). As repetition of the disaccharide N-acetylglicosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) is the only common part in A- and B-type PG, LysM domains most likely bind to this component of the PG. No binding to (GlcNAc)\textsubscript{3} was obtained with the LysM domain of AcmA (A. Steen et al., unpublished).

\[ \text{\textbf{Fig. 1. Taxonomic coverage of the LysM domain. Picture was generated by the interpro website (http://www.ebi.ac.uk/interpro/}} \]

\[ \text{\textbf{Fig. 2. Consensus sequence of LysM (Pfam database entry PF01476) (adapted from Desvaux et al., 2006). Above the sequence the location and length (N = number of amino acid residues) of the inserts among the LysMs are indicated. The consensus residues of the YG domain (YXXXGXXH, Turner et al., 2004) that are also present in the GW domain (Prodom entry PD006903) and the Cholin domain (PD000439) are boxed. Boxes below the sequence indicate regions with secondary structure as defined by Mulder et al. (2006).}} \]
Members of the type III sugar binding antiviral protein Cyanovirin-N Homology (CVNH) contain one LysM. The CVN domain is a small 11 kDa domain that binds to the mannose moieties of surface glycoproteins of viruses like HIV and Ebola, thereby blocking entry of the virus into the host (Percudani et al., 2005). Type III proteins of the CVN homology family are only present in filamentous ascomycetes and contain two CVNH domains separated by a LysM. Interestingly, β1,4-N-acetylglucosamine is the reducing end of the Man9 oligosaccharide moiety associated with the viral surface glycoproteins gp120.

In some eukaryotes, e.g. in the nematode Caenorhabditis elegans and in the green alga Volvox carteri, LysMs are present in chitinases (Bateman and Bycroft, 2000), suggesting that they may bind to chitin, a polymer of GlcNAc that resembles the glycan chain of PG. The striking sequence differences between LysMs of chitinases and those of the bacterial cell wall-hydrolysing enzymes, in particular the presence of multiple Cys residues in the former, could account for the difference in binding specificity. The Nod factors secreted by Rhizobium species, recognized by the LysMs domains of plant receptors, consist of a GlcNAc backbone of four or five residues (Radutoiu et al., 2007). Chitin-like compounds are not only involved in the signalling between bacteria and plants, they are also involved in inducing plant defence responses against fungi, which have chitin as the major component of their cell walls. In Arabidopsis and rice the defence response against fungi is induced via the LysM-containing receptor-like kinase CERK1, LysM RLK1 and CEBiP respectively (Kaku et al., 2006; Miya et al., 2007; Wan et al., 2008).

Recently it has been proven that the two N-terminal LysMs of the chitinase from Pteris ryukyuensis bind (GlcNAc)5 with a stoichiometry of 1:1 (Ohnuma et al., 2008). Poor binding was obtained with (GlcNAc)3 while an increasingly stronger binding was obtained with (GlcNAc)4 and (GlcNAc)5.

Together these data show that GlcNAc is the common sugar bound by LysMs, but binding studies using LysMs from different proteins and a variety of possible substrates are needed to determine whether besides GlcNAc, MurNAc, for example, possibly together with some of the peptide stem residues in PG, is also recognized by LysMs.

**Structure of the Lysin motif**

The structure of the LysM of the Escherichia coli outer membrane-bound lytic murein transglycosylase MltD has been solved by NMR (Bateman and Bycroft, 2000). The LysM has a βαβαβ secondary structure with the two α-helices packing onto the same side of an antiparallel β-sheet. Using this information, the structure of the CVNH-LysM protein from a filamentous ascomycete was predicted (Percudani et al., 2005). The crystal structure of the Bacillus subtilis spore protein YkuD, which contains one N-terminally located LysM, was solved by multi-wavelength anomalous dispersion (Bielnicki et al., 2006). The structure of the LysM in YkuD was similar to that of MltD. Two glycine residues (residues 6 and 38 in Fig. 2) are highly conserved in LysMs and are part of tight turns in the LysM structure. The highly conserved asparagine at position 28 in Fig. 2 forms a turn at the end of helix 2 of the LysM (Bateman and Bycroft, 2000).

The structures of the three LysMs in the receptor-like kinase NFP of the legume Medicago trunculata, using hydrophobic cluster analysis plots, showed that the first LysM is positively charged or neutral, but that the second two motifs are predominantly electronegative (Arrighi et al., 2006). Nodulating plants, like M. trunculata and *...*
Lotus japonicus, are able to recognize specific Nod factors, which are secreted by their symbiotic bacteria. Leucine 118 of the second LysM of NFR5 of L. japonicus is the main determinant in the recognition of the Nod factor of Rhizobium leguminosarum biovar viciae strain DZL. This Leu-118 corresponds to residue 5 in Fig. 2. The DZL Nod factor is not recognized by Lotus filicaulis NFR5, which has a lysine at position 118 (Radutoiu et al., 2007). Structural studies should elucidate how Leu-118 determines this specificity.

Mulder et al. (2006) also describe the homology modelling of the three LysMs of M. trunculata using the MitD LysM structure. They docked chito-oligosacharides and Nod factors on these models to predict the most favoured binding modes of these compounds. The Nod factors specific for M. trunculata are sulphated and O-acetylated and seem to bind in similar orientation to all three LysMs. The sulphate group of the Nod factor is in close proximity to a Lys residue, allowing a salt bridge between the two. The O-acetyl group of the GlcNAc in the Nod factor is located in the hydrophobic pocket of the LysM, made mainly by amino acids of the loop between strand beta1 and helix alpha1. The lipid moiety of the Nod factor wraps around the LysM and seems to interact with amino acids that are at the base of beta2. The M. trunculata LysMs are highly glycosylated but the glycosylation does not interfere with Nod factor binding. Ohnuma et al. (2008) characterized the carbohydrate binding site of the two LysMs of the chitinase of P. ryukyuensis using calorimetric and NMR techniques. They identified residues critical for (GlcNAc)₃ binding by titrating the (GlcNAc)₃ and monitoring by NMR. The N-terminal part of helix 1 and the loop between strand 1 and helix 1, together with the C-terminal part of helix 2 and the loop between helix 2 and strand 2, form a shallow groove by a cluster of hydrophobic residues. Moreover, a Tyr residue (at position 10 in Fig. 2) is of critical importance in (GlcNAc)₃ binding.

It will be a challenge to elucidate the structure and specific interactions of a LysM (protein) bound to its substrate, especially one containing multiple LysMs.

Multiple functions of LysM-containing proteins

Peptidoglycan hydrolases

Most of the LysM-containing proteins are bacterial PG hydrolases (Fig. 1). Interestingly, LysM domains of glycosylases, such as N-acetylmuramidases and N-acetylglucosaminidases, are located downstream of the active-site domain at the C-terminus of the protein. In PG endopeptidases, such as those of the Cysteine, Histidine-dependent Amidohydrolases/Peptidases (CHAP) superfamily (Layec et al., 2008), they are present in the N-terminal part upstream of the active site (Table 1). A possible reason for the different topology could be proper positioning of the active-site domains towards their specific substrates.

The number of LysMs in PG hydrolases and their location in the proteins can vary considerably (Table 1 and Table S1). The presence of three C-terminal LysMs in the LysM domain of the N-acetylglucosaminidase AcmA of L. lactis appears to be optimal for lactococcal cell wall degradation (Steen et al., 2005a); AcmA derivatives containing one, two or four LysMs bound less efficiently to L. lactis cells, resulting in reduced cell lysis. When all LysMs were removed, PG binding was lost, concomitant with an almost total loss of PG hydrolysing activity.

### Table 1. Overview of the number (nr) and the relation between the N- or C-terminal (term) location of LysMs and the specificity of peptidoglycan hydrolases.

<table>
<thead>
<tr>
<th>Specificity</th>
<th>Pfam</th>
<th>Location</th>
<th>N-term (nr)</th>
<th>C-term (nr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mannosyl-glycoprotein endo-beta-N-acetylglicosaminidase</td>
<td>PF01832</td>
<td>Sec</td>
<td>3</td>
<td>1 till 6</td>
</tr>
<tr>
<td>N-acetylmuramoyl-L-alanine amidase</td>
<td>PF01520</td>
<td>Sec</td>
<td>5</td>
<td>1 till 4, 8 or 9</td>
</tr>
<tr>
<td>Phage lysozyme</td>
<td>PF00959</td>
<td>Sec</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Glycosyl hydrolases family 18</td>
<td>PF00704</td>
<td>Sec</td>
<td>1 till 4, or 12</td>
<td></td>
</tr>
<tr>
<td>Membrane-bound lytic murein transglycosylase D</td>
<td>PF06474</td>
<td>Lipo</td>
<td>1 till 4, 8</td>
<td></td>
</tr>
<tr>
<td>Transglycosylase SLT domain</td>
<td>PF01464</td>
<td>Mem</td>
<td>3</td>
<td>1 till 8</td>
</tr>
<tr>
<td>Glycosyl hydrolases family 25</td>
<td>PF01183</td>
<td>Sec</td>
<td>1 till 3</td>
<td></td>
</tr>
<tr>
<td>Plant self-incompatibility protein S1</td>
<td>PF05993</td>
<td>4 or 12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NlpC/P60 family</td>
<td>PF00877</td>
<td>Sec/Lipo</td>
<td>1 till 5</td>
<td></td>
</tr>
<tr>
<td>Cysteine, Histidine-dependent Amidohydrolases/Peptidases (CHAP)</td>
<td>PF05257</td>
<td>Sec</td>
<td>1 till 2</td>
<td></td>
</tr>
<tr>
<td>Peptidase family M23</td>
<td>PF01551</td>
<td>Sec/Lipo/Mem</td>
<td>1 till 4</td>
<td></td>
</tr>
<tr>
<td>Zinc carboxypeptidase</td>
<td>PF00246</td>
<td>Sec</td>
<td>1 till 2</td>
<td></td>
</tr>
<tr>
<td>EnfK/YbiS/YctS/YnhG</td>
<td>PF03734</td>
<td>Sec/Lipo/Mem</td>
<td>1 till 3</td>
<td></td>
</tr>
<tr>
<td>Bacterial Ig-like domain (group 1)</td>
<td>PF02369</td>
<td>Mem</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Bacterial Ig-like domain (group 2)</td>
<td>PF02368</td>
<td>Sec/Mem</td>
<td>1 or 2</td>
<td></td>
</tr>
</tbody>
</table>

a. Sec, secreted protein; Mem, membrane protein; Lipo, lipoprotein. In grey the groups containing the highest number of proteins with a certain specificity are indicated.

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd, Molecular Microbiology, 68, 838–847
Similarly, removal of the C-terminal six LysMs from the orthologous N-acetylglucosaminidase AtlA of *E. faecalis* led to a 580-fold reduction of enzyme activity (Eckert *et al.*, 2006). Whether the number of motifs is optimized for a specific substrate is unknown.

AcmA is subject to proteolytic degradation by the extracellular proteases PrtP and HtrA (Buist *et al.*, 1998; Poquet *et al.*, 2000; Steen *et al.*, 2005b). Most likely, the protein is cleaved within the Ser/Thr/Asp-rich intervening sequences between the LysMs. The consequent reduction of the number of LysMs results in a reduced enzymatic activity of AcmA. Proteolytic degradation of LysMs from PG hydrolases is generally observed (Steen *et al.*, 2005b; Eckert *et al.*, 2006; Fukushima *et al.*, 2006); whether this degradation within the LysM domain is a directed manner to control the potentially lethal cell wall degrading activity is an interesting but as yet unproven strategy.

Actinobacteria can enter a state in which they are less metabolically active and lose culturability, a process that is controlled by resuscitation-promoting factors (RPFs). Rpf of *Micrococcus luteus* is essential: it has been shown to possess muramyl activity that is probably responsible for its observed activity in resuscitation (Mukamolova *et al.*, 2006). Rpf belongs to the LysM subfamily of the firmicute Sps proteins, of which all members contain one C-terminal LysM. Proteins of the SpsA subfamily contain two N-terminal LysMs (Ravagnani *et al.*, 2005). Proteins of both subfamilies have been suggested to be secreted, muramyl enzymes that control bacterial culturability via enzymatic modification of the bacterial cell envelope.

**Phage lysins**

LysM domains are present in bacteriophage lysins (muramidases) such as those from *L. lactis* phage Tuc 2009 and the *B. subtilis* phages phi 29, PBSX and PZA (Buist *et al.*, 1995). Phage PBSX encodes the lysins XkdP (specificity unknown), XlyA and XlyB (both amidases), which each contain one LysM. The LysMs in XlyA and XlyB are combined with the domain PG-binding_1. The presence of this additional substrate-binding domain suggests that each protein binds to two different regions of the PG substrate to properly position the active-site domain(s).

**Virulence factors**

Several LysM domain-containing proteins are virulence factors of human bacterial pathogens. *Staphylococcus aureus* produces five LysM proteins which are all involved in virulence (Table S1). The mature N-acetylmuramyl-L-alanine amidase Sle1 of *S. aureus* contains three N-terminal LysMs and is involved in cell separation (Kajimura *et al.*, 2005). Cells of a sle1 mutant form clusters and the strain is significantly attenuated in an acute-infection mouse model, indicating that cluster formation affects spread of the bacteria during infection. The *S. aureus* membrane protein EbpS contains an N-terminal elastin-binding domain for binding to tissue cells, which is exposed to the cell surface and one C-terminal LysM, which is not exposed on the cell surface and was suggested to be buried within the PG (Fig. 3) (Downer *et al.*, 2002). The covalently PG-bound protein A of *S. aureus* also contains a LysM immediately upstream of the LPxTG box (Bateman and Bycroft, 2000), suggesting an additional binding and/or positioning mode (Fig. 3). *Listeria monocytogenes* expresses six LysM proteins (Bierne and Cossart, 2007). The P60 PG hydrolase contains an N-terminal and a central LysM, which are separated by an SH3 domain that has a putative PG binding function. Mutations in *p60* and *namA* (three C-terminal LysMs motifs) resulted in intermediate reduction of bacterial virulence *in vivo* (Lenz *et al.*, 2003). The authors speculated that the combined action of the two PG hydrolases results in the production of glucosaminylmuramyl dipeptide, which is known to modify host inflammatory responses. The surface immunogenic protein (Sip) of *Streptococcus agalactiae*, which is conserved in strains of group B *Streptococcus* (GBS), contains one N-terminal LysM (Borges *et al.*, 2005). Administration of Sip-specific antibodies to pregnant mice resulted in protection of the pups against a lethal challenge with GBS (Martin *et al.*, 2002). Sip was only protective against GBS strains without a polysaccharide capsule, showing that accessibility of Sip, which is exposed on the cell surface at the polar sites and the septal region (Rioux *et al.*, 2001), can be blocked by such capsules. The LysM domain-containing proteins FsaP, TspA and Intimin of, respectively, *Francisella tularensis* (Mellilo *et al.*, 2006), *Neisseria meningitidis* (Oldfield *et al.*, 2007) and enterohaemorrhagic *E. coli* (Bateman and Bycroft, 2000) are involved in adherence to human cells. All three proteins are located in the outer membrane (Fig. 3). The mature proteins have similar modular structures, in which a centrally located transmembrane domain is preceded by a LysM domain, suggesting that the N-terminal part of the proteins is in contact with the PG in the periplasm and the C-terminal part is exposed on the cell surface. It is unknown why these proteins interact with the PG. A virulent strain of *F. tularensis* expresses elevated levels of FsaP and *F. tularensis*-infected mice developed antibodies against FsaP. Mutation of *tspA* in *N. meningitidis* resulted in reduced binding to Hsp-2 and meningothelial monolayers (Melillo *et al.*, 2006).
**LysM domain-containing proteins of eukaryotes**

LysM domains are also present in (putative) proteins of plants, fungi and animals, including man. Only some of the LysM proteins have been studied, e.g. the aforementioned chitinases of *C. elegans*, but the function of most of these proteins is still unknown. As mentioned before the chitinase of *C. elegans* has several LysMs that contain consensus Cys residues. The presence of Cys residues suggests that disulphide bridges are used to maintain protein structure of the motif (Bateman and Bycroft, 2000), or alternatively could allow for cofactor binding.

The legume *L. japonicus* has two LysM-type serine/threonine receptor kinases, NFR1 and NFR5, which enable the plant to recognize its bacterial symbiont *Mesorhizobium loti* (Fig. 3) (Madsen et al., 2003; Radutoiu et al., 2003). *M. trunculata* contains seven LysM domain-containing receptor-like kinases. The authors speculate that also in this plant species the LysM domains are involved in recognition of the bacterial symbiont (Limpens et al., 2003). The symbiotic bacteria secrete Nod factors, which resemble chitin and are thought to bind to the LysM domains of the plant receptors. Eventually, this would lead to endocytosis of the bacteria by plant cells and subsequent nodulation in the plant roots. Besides these receptor kinases involved in symbiotic signalling, plant LysM proteins are also involved in defence signalling against fungal attack. Chitin is involved in inducing defence responses in both monocots and dicots. Recently it has been shown that CEBiP of *Oryza sativa* contains two LysMs and is a glycoprotein with high affinity for chitin oligosaccharides. Knock-down of CEBiP resulted in suppression of chitin-induced defence response (Kaku et al., 2006). Arabidopsis CERK1 is a receptor-like kinase with three LysMs that also plays a key role in fungal perception by the plant (Miya et al., 2007).

**Localized cellular binding of LysM domains**

Although PG covers the entire bacterial cell surface, immunofluorescence studies showed that the C-terminal LysM domain of AcmA binds to specific sites on the bacterial surface, i.e. near the poles and septum of *L. lactis* cells (Steen et al., 2003). Interestingly, our recent results show that protein secretion also takes place near the septum (Buist et al., 2006). The secondary polymers lipoteichoic acids (LTAs), which consist of polyglycerol-phosphate, are responsible for the specific binding of AcmA: AcmA does not bind where LTAs are present (Steen et al., 2003). Also the presence of surface layer proteins hinders the binding of the LysM domain of AcmA to *Lactobacillus helveticus* (Steen et al., 2003) and, thus, localization of binding was found to be species-specific. Increased O-acetylation of the PG of *L. lactis* results in resistance of the cells to AcmA, possibly by reduced binding of AcmA (Veiga et al., 2007), while de-N-acetylation of lactococcal PG did not affect the binding of AcmA (Meyrand et al., 2007). Interestingly, Aaa of *S. aureus*, which has three N-terminal LysMs, seems to bind all over the cell surface (Heilmann et al., 2005).

The LysM domain of Sip of *S. agalactiae* is responsible for localized binding of the protein at the cell poles and the septal region (Rioux et al., 2001). Polar and septal binding has also been shown for the D,L-endopeptidases LytE, CwlS and LytF of the rod-shaped bacterium *B. subtilis*. These proteins contain three, four and five N-terminal LysMs respectively (Yamamoto et al., 2003; Fukushima et al., 2006). LytE, prior to secretion, interacts with membrane-associated MreBH, which is required for the helical pattern of extracellular localization of LytE into the cylindric part of the cell wall (Carballido-Lopez et al., 2006). The pattern of LytE localization resembles that of the SecA and SecY proteins of the general Sec-secretion machinery of *B. subtilis* (Campo et al., 2004). No interaction of MreBH with the latter two proteins was found. Whether localization of secretion of the LysM proteins coincides with their location of activity is also something that needs further study.

The PG hydrolase YneA of *B. subtilis* is responsible for suppression of cell division during SOS response. YneA consists of only 105 aa and has a putative N-terminal transmembrane domain that is separated by only 10 aa from a centrally located LysM of Gram-positive bacteria with a pI below 7. Recently, it has been shown that the legume *L. lactis* and *B. subtilis* contain a short transmembrane segment (Carballido-Lopez et al., 2003) and, thus, de-N-acetylation of lactococcal PG did not affect the binding of AcmA (Meyrand et al., 2007), while de-N-acetylation of lactococcal PG did not affect the binding of AcmA (Meyrand et al., 2007). Interestingly, Aaa of *S. aureus*, which has three N-terminal LysMs, seems to bind all over the cell surface (Heilmann et al., 2005).

The LysM domain of Sip of *S. agalactiae* is responsible for localized binding of the protein at the cell poles and the septal region (Rioux et al., 2001). Polar and septal binding has also been shown for the D,L-endopeptidases LytE, CwlS and LytF of the rod-shaped bacterium *B. subtilis*. These proteins contain three, four and five N-terminal LysMs respectively (Yamamoto et al., 2003; Fukushima et al., 2006). LytE, prior to secretion, interacts with membrane-associated MreBH, which is required for the helical pattern of extracellular localization of LytE into the cylindric part of the cell wall (Carballido-Lopez et al., 2006). The pattern of LytE localization resembles that of the SecA and SecY proteins of the general Sec-secretion machinery of *B. subtilis* (Campo et al., 2004). No interaction of MreBH with the latter two proteins was found. Whether localization of secretion of the LysM proteins coincides with their location of activity is also something that needs further study.

The PG hydrolase YneA of *B. subtilis* is responsible for suppression of cell division during SOS response. YneA consists of only 105 aa and has a putative N-terminal transmembrane domain that is separated by only 10 aa from a centrally located LysM of Gram-positive bacteria with a pI below 7. Recently, it has been shown that the legume *L. lactis* and *B. subtilis* contain a short transmembrane segment (Carballido-Lopez et al., 2003) and, thus, de-N-acetylation of lactococcal PG did not affect the binding of AcmA (Meyrand et al., 2007), while de-N-acetylation of lactococcal PG did not affect the binding of AcmA (Meyrand et al., 2007). Interestingly, Aaa of *S. aureus*, which has three N-terminal LysMs, seems to bind all over the cell surface (Heilmann et al., 2005).

The LysM domain of Sip of *S. agalactiae* is responsible for localized binding of the protein at the cell poles and the septal region (Rioux et al., 2001). Polar and septal binding has also been shown for the D,L-endopeptidases LytE, CwlS and LytF of the rod-shaped bacterium *B. subtilis*. These proteins contain three, four and five N-terminal LysMs respectively (Yamamoto et al., 2003; Fukushima et al., 2006). LytE, prior to secretion, interacts with membrane-associated MreBH, which is required for the helical pattern of extracellular localization of LytE into the cylindric part of the cell wall (Carballido-Lopez et al., 2006). The pattern of LytE localization resembles that of the SecA and SecY proteins of the general Sec-secretion machinery of *B. subtilis* (Campo et al., 2004). No interaction of MreBH with the latter two proteins was found. Whether localization of secretion of the LysM proteins coincides with their location of activity is also something that needs further study.
B. subtilis (Chauhan et al., 2006). Rv2719c has cell wall hydrolase activity, is a potential regulator of M. tuberculosis cell division and the amount of the protein and, possibly, its activity are modulated under a variety of growth conditions. Localization of GFP–Rv2719c fusion protein led to the suggestion that Rv2719c is targeted to potential PG synthesis zones, i.e. to the poles and to sites close to mid-cell.

LysM domains also play a role in the development of spores in sporulating bacteria. When the signal peptide of E. coli β-lactamase was replaced by the two LysMs of the B. subtilis spore protein YaaH, the fusion protein was directed to the surface of the developing spore. Apparently, the LysM domains of spore proteins serve as localization signals to forespores. SafA and SpoVID localize to the cortex–coat interface encircles the entire cortex layer of the spore (Fig. 3) (Costa et al., 2006). SafA is initially targeted independently of SpoVID, possibly via its LysM domain, while in a second stage SafA forms a complex with SpoVID via direct protein–protein interactions. Both proteins promote attachment of the spore coat to the cortex and both proteins and their complex are likely to interact with additional coat components (Costa et al., 2006). Interestingly SafA and SpoVID have a rather low pl; whether this is important for localized binding in the spore coat remains to be answered. Localized and controlled binding of spore proteins to the cortex–coat interface via LysM domains, as a general concept to induce protein complex formation in the spore coat, is a very interesting model but remains to be further investigated.

Together these findings show that proteins containing a LysM domain bind PG at specific loci depending on the presence of secondary cell wall polymers and, possibly, PG modifications.

### Evolution of LysM domains

The presence of LysMs in many bacterial species as well as many eukaryotes, and the absence of the motif in archaeal proteins raises questions about their origin and evolution. Two theories can be envisaged: LysM was present in the common ancestor of all life, but was lost in the archael lineage, or it evolved after the separation of bacteria, archaea and eukaryotes and was subsequently transferred from bacteria to eukaryotes, or vice versa (Ponting et al., 1999). Because of their omnipresence in the latter kingdoms, the second theory requires that the transfer occurred early in evolution, or later via multiple transfers. Phylogenetic information is, however, not sufficient to resolve the earliest relationships between LysM in bacteria and eukaryotes. In bacteria, the LysM seems to have evolved into a general PG-binding motif, whereas in eukaryotes it is a chitin-binding motif.

Motif similarities suggest that plant LysMs are ancient and have evolved through local and segmental duplications. The family has undergone further duplication and diversification in legumes, where some LysM kinases function as receptors for bacterial nodulation factors. Plant LysMs eventually evolved into a minimum of six distinct types in LysM kinases and five additional types in non-kinase proteins (Zhang et al., 2007).

Alignment of bacterial LysM domains with two other carbohydrate-binding motifs in proteins revealed a shared YG motif (Turner et al., 2004). The cholin-binding domains of cell surface proteins binding to cholin in LTA of Streptococcus pneumoniae and Clostridia species, as well as the LTA-binding GW domain of LTA-binding proteins of Listeria species, also contain the YG motif (Fig. 2). It is possible that an ancestral carbohydrate-binding domain in Gram-positive bacteria has evolved into LysM, choline binding and GW domains. Sugar-binding proteins usually have aromatic residues, which enable contact between the protein surface and the sugar. The YG motif could therefore also have been the result of parallel evolution.

### Concluding remarks

The LysM domain is a widely spread domain in nature that has been shown to bind different PG types in bacteria, chitin-like compounds in eukaryotes and a viral glycoprotein. GlcNAc seems to be an important constituent of the LysM ligand and has been shown to interact with LysMs of some eukaryotic proteins. Whether GlcNAc is the sole moiety recognized by LysM remains to be elucidated. In bacterial PG hydrolases, the motif is possibly needed to properly position the active-site domain(s) towards their substrate (Steen et al., 2005a). In legume plants, LysM plays a crucial role in establishing the interaction of the plant with nitrogen-fixing bacteria by binding Nod factors secreted by the bacteria. In other plants, a defence response against fungi is induced via binding of GlcNAc to LysM-containing receptor-like kinases. For other LysM-containing proteins the nature of ligand still has to be established, for both the entire LysM protein and its individual LysM(s). Especially, the LysM domains of animal and fungal proteins need attention.

The structure of only two LysMs is known and more structures need to be determined. Until then, modelling of LysM structures is helpful to establish how and where the ligand binds. Eventually, crystal or NMR structures of prokaryotic and eukaryotic LysMs together with their specific ligands will be essential to fully understand ligand binding.

LysM is often found in multiple copies in proteins and the exact function of these duplications also needs attention, as only for the L. lactis autolysin AcmA has it been
shown that three LysMs are optimal for proper enzyme functioning. Characteristics of the LysM domains such as their pl, glycosylation, presence of disulfide bridges, aa insertions compared with the consensus sequence (Fig. 2) and composition and length of the intervening sequences between the LysMs need to be investigated.

More is known about the subcellular localization of LysM proteins in bacteria, an important aspect as, e.g. in the case of AcmA, this process seems to regulate the potentially lethal activity of the autolysin. The subcellular localization of the proteins is directed by their LysM domains and further influenced by cell wall components such as other proteins, LTA and modifications of the PG. Future research should address the (combined) effects of these cell wall components on localization at the molecular level.

Finally, as most of the LysM proteins are secreted and bind PG, it will be interesting to establish the impact of possible intercellular action, such as binding of the proteins to other cells, for instance, for in trans cell lysis between species in mixed cultures, aggregation and biofilm formation. An example of biotechnological application of LysMs is the use of the LysM domain of AcmA to bind antigens to other cells, for instance, for subcutaneous immunizations of mice with the immobilized recombinant antigen–LysM fusion proteins could subsequently bind to non-genetically modified Gram-positive bacteria for oral immunization purposes (Bosma et al., 2006; van Rossmalen et al., 2006). To that end, the LysM domain was fused to the C-terminus of antigens of the malaria parasite Plasmodium falciparum or S. pneumoniae, and produced and secreted by L. lactis. The secreted antigen–LysM fusion proteins could subsequently bind to non-genetically modified bacteria or their purified cell walls. After oral or subcutaneous immunizations of mice with the immobilized recombinant antigens, a specific, possibly protective, antibody response was obtained. Many more applications, e.g. in modulating host–microbe interactions and surface display of enzymes based on the specific binding properties of the LysMs, are foreseen.

References


endopeptidase gene product, YojL (renamed CwlS), plays a role in cell separation with LytE and LytF in *Bacillus subtilis*. 


Garvey, K.J., Saedi, M.S., and Ito, J. (1986) Nucleotide sequence of *Bacillus* phage phi 29 genes 14 and 15: homology of gene 15 with other phage lysozymes. 


*Proc Natl Acad Sci USA* **103**: 11086–11091.


*Proc Natl Acad Sci USA* **100**: 12432–12437.


*Microbiology* **153**: 3275–3285.


*Proc Natl Acad Sci USA* **104**: 19613–19618.


*Glycobiology* **16**: 801–809.


*Proteins* **60**: 670–678.


*EMBO J* **26**: 3923–3935.


*BMC Genomics* **6**: 39.


*Methods* **38**: 144–149.


**LysM, a widely distributed glycan binding domain**


Supplementary material

1This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-2958.2008.06211.x

(This link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

---

*Lactococcus lactis* is an *N*-acetylglucosaminidase with an optimal number of LysM domains for proper functioning. *FEBS J* **272**: 2854–2868.


