Research

Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes

Vivek Anantharaman and L Aravind

Address: National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.

Correspondence: L Aravind. E-mail: aravind@ncbi.nlm.nih.gov

Published: 3 February 2003

Genome ***Biology*** 2003, **4**:R11

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2003/4/2/R11

Received: 9 October 2002
Revised: 3 December 2002
Accepted: 20 December 2002

Abstract

Background: Peptidoglycan is hydrolyzed by a diverse set of enzymes during bacterial growth, development and cell division. The NlpC/P60 proteins define a family of cell-wall peptidases that are widely represented in various bacterial lineages. Currently characterized members are known to hydrolyze D-γ-glutamyl-meso-diaminopimelate or *N*-acetylmuramate-L-alanine linkages.

Results: Detailed analysis of the NlpC/P60 peptidases showed that these proteins define a large superfamily encompassing several diverse groups of proteins. In addition to the well characterized P60-like proteins, this superfamily includes the AcmB/LytN and YaeF/YiiX families of bacterial proteins, the amidase domain of bacterial and kinetoplastid glutathionylspermidine synthases (GSPSs), and several proteins from eukaryotes, phages, poxviruses, positive-strand RNA viruses, and certain archaea. The eukaryotic members include lecithin retinol acyltransferase (LRAT), nematode developmental regulator Egl-26, and candidate tumor suppressor H-rev107. These eukaryotic proteins, along with the bacterial YaeF/poxviral G6R family, show a circular permutation of the catalytic domain. We identified three conserved residues, namely a cysteine, a histidine and a polar residue, that are involved in the catalytic activities of this superfamily. Evolutionary analysis of this superfamily shows that it comprises four major families, with diverse domain architectures in each of them.

Conclusions: Several related, but distinct, catalytic activities, such as murein degradation, acyl transfer and amide hydrolysis, have emerged in the NlpC/P60 superfamily. The three conserved catalytic residues of this superfamily are shown to be equivalent to the catalytic triad of the papain-like thiol peptidases. The predicted structural features indicate that the NlpC/P60 enzymes contain a fold similar to the papain-like peptidases, transglutaminases and arylamine acetyltransferases.

Background

The rigid cell wall that forms a protective layer around most bacterial cells is chiefly composed of peptidoglycan, a biopolymer unique to bacteria [1,2]. The backbone of peptidoglycan consists of a chain of alternating *N*-acetylglucosamine (NAG) and *N*-acetylmuramate (NAM) units linked by a 1-4 glycosidic bond between the two hexoses. The NAM units of the glycan chain are linked to short peptides, which are synthesized via a ribosome-independent mechanism, and contain both canonical L-amino acids and unusual D-amino acids. Cross-links between these peptides hold together the glycan chains and give the cell wall its characteristic rigidity.

While this generic structure is conserved throughout the Bacteria there are number of variations, in particular lineages of bacteria, in terms of adducts to the glycan, and composition of the peptide chain [1,2]. A large suite of conserved enzymes, such as glycosyltransferases (which form the hexose polymers), racemases (which generate D-amino-acid units), and peptidyltransferases and transpeptidases (which form interpeptide linkages) are involved in the biosynthesis of peptidoglycan [3-5]. The bacterial cell wall is dynamic, and undergoes reorganization during vegetative growth, development and cell division [1,2,6-8]. In these processes the wall is disassembled through the action of a diverse set of enzymes that hydrolyze various linkages in peptidoglycan. These enzymes include glycosidases, such as lysozymes, that attack the polysaccharide backbone, and peptidases that degrade the cross-linking peptides [8]. Some of these enzymes are also encoded by bacteriophages and are used to degrade the host cell wall [9-11].

Biochemical and structural analysis of the peptide-hydrolyzing enzymes of bacteria have revealed a large diversity of peptidases with different catalytic mechanisms. Some well studied examples of these peptidases include Zn-dependent peptidases of the Hedgehog carboxy-terminal domain fold and the DD-peptidases with a PAS domain-like fold [12-17]. The *Bacillus subtilis* autolysins LytE and LytF, the *Listeria monocytogenes* proteins p60 and p45 and the *Streptococcus* PcsB proteins define another family of somewhat less well explored peptidoglycan-hydrolyzing proteins that are present in all bacterial lineages [8,18-21]. This family includes the *Escherichia coli* membrane-associated lipoprotein NlpC, and possesses a catalytic domain containing a conserved amino-terminal cysteine and a carboxy-terminal histidine. Members of this family (here termed the NlpC/P60 family) from the genus *Bacillus* have been shown to be D,L endopeptidases that hydrolyze the D-γ-glutamyl-meso-diaminopimelate linkage in the cell-wall peptides [19]. In *B. subtilis*, disruption of LytE (CwlF) and LytF (CwlE) causes the normally rod-shaped cells to assume a long filament-like morphology as a result of defective cell-wall separation during cell division [18,19,22]. In *L. monocytogenes*, *p60* is an essential gene and has a similar role in cell separation [23]. Multiple paralogous proteins with NlpC/P60 catalytic domains are present in most bacteria, suggesting that the NlpC/P60 family is a peptidase family with a widespread role in the dynamics of the bacterial cell wall.

We sought to understand the evolutionary history of the proteins involved in cell-wall dynamics, because such an analysis could throw light on the emergence and diversification of this uniquely bacterial structure. Such an investigation could also uncover hitherto under-appreciated components of cell-wall dynamic systems and could help in clarifying the biochemical mechanisms of certain poorly characterized but critical proteins. Furthermore, there is considerable interest in these proteins as they could be potential targets for antibacterial agents, and some cell-wall-hydrolyzing enzymes from bacteriophages themselves function as antibacterial agents. As a part of a comprehensive analysis of proteins involved in bacterial cell-wall dynamics we studied the diversity of the NlpC/P60 family. Here we present evidence that these proteins define a large superfamily, which encompasses several conserved lineages of proteins which were previously not known to be related to the NlpC/P60 proteins. These include the cell-wall hydrolases typified by the Pal protein from streptococcal phages [24,25] and the amidase domain of the bacterial and kinetoplastid GSPS [26].

We also show that members of the NlpC superfamily exist outside the bacterial superkingdom, in eukaryotes, large DNA viruses, positive-strand RNA viruses and certain archaea. In eukaryotes, one of the members of this family has been studied experimentally, and possesses LRAT activity rather than peptidase activity [27]. These eukaryotic versions, along with certain bacterial forms, show a circular permutation of the domain which results in a swapping of the positions of the catalytic cysteine and histidine residues in the sequence. These observations point to greater diversity in terms of biochemistry, biological functions and structural organization than has previously been recognized in this superfamily.

Results and discussion

Sequence analysis and detection of novel divergent members

To determine the entire extant of the NlpC/P60 family we initiated PSI-BLAST [28] searches of the non-redundant database (expect (E)-value for inclusion in threshold = 0.01, iterated to convergence) with a number of starting points, such as the catalytic domains of *B. subtilis* LytF (ClwE) and LytE (ClwF), *L. monocytogenes* P60 and P45 and *E. coli* NlpC. A search with the NlpC of *E. coli* (gi:15802120, residues 35-154) recovered several P60 homologs in its first iteration, YaeF of *E. coli* (second iteration, $E = 3 \times 10^{-4}$), the *Arabidopsis* homolog of LRAT, At5g16360 (gi:15237330, second iteration, $E = 5 \times 10^{-4}$) and LytN of *Staphylococcus aureus* (gi:3767593, fourth iteration, 8×10^{-6}). Reciprocal searches initiated with regions from the newly detected proteins YaeF, LRAT and LytN detected their closely related homologs in the initial iterations, but in subsequent iterations they recovered several classic NlpC/P60 homolog members with significant E-values ($E < 0.01$), suggesting that these proteins defined a novel superfamily of proteins.

The large set of homologous proteins from diverse bacteria which was recovered in these searches typically contain a characteristic amino-terminal conserved region with a cysteine, followed by a carboxy-terminal conserved region with a characteristic histidine (Figure 1; see also Additional data files). In between these conserved regions there is another

characteristic stretch with a highly conserved glycine that is often followed by an aspartate (Figure 1). Hereinafter, we term these obviously related bacterial proteins as the classical NlpC/P60 group. Thiol peptidases of the papain-like, transglutaminase-like, adenoviral protease-like, and caspase-hemoglobinase-like folds have a conserved cysteine (C)-histidine (H) pair [29-32] (Figure 2). However, the context of the C-H pair in the NlpC/P60 peptidase domains, along with the central motif with the conserved glycine (G), defines a unique constellation of residues (Figure 1) which is not encountered in any of the above peptidase families. Accordingly, this pattern served as a good marker to identify divergent homologs of the classical NlpC/P60 family. The above searches unexpectedly recovered LRAT and its homologs from various eukaryotes, *Caenorhabditis elegans* developmental regulator Egl-26, candidate tumor suppressor H-rev107, proteins from poxviruses and animal positive-strand RNA viruses with statistically significant E-values. Transitive searches initiated with the divergent members detected in these searches additionally recovered the amidase domains of bacterial and kinetoplastid GSPS and the bacterial peptidoglycan-degrading enzymes such as *Lactococcus* AcmB, *Staphylococcus* LytA [33] and LytN [34] and their homologs.

The presence of three tandem repeats of the NlpC/P60 module in the protein YwtD from *B. subtilis* allowed us to define the amino- and carboxy-terminal boundaries of this domain quite precisely. The domain begins approximately 25-30 residues before the amino-terminal conserved cysteine, and ends 40-55 residues from the carboxy-terminal conserved histidine. A search for conserved motifs using the Gibbs sampling procedure resulted in the detection of three conserved motifs with a probability of chance occurrence < 10^{-12} in the search space comprising these proteins. These motifs correspond to the above-mentioned conserved signatures, which are characteristic of the NlpC/P60-like proteins, and were readily identifiable in all true positives recovered in the transitive sequence profile searches. Interestingly, the eukaryotic versions typified by the LRAT, their viral homologs and the bacterial forms typified by *E. coli* YaeF, while containing all the hallmark signatures of the NlpC/P60 catalytic domain, showed a reversal in the positions of the motifs with the conserved cysteine and the conserved histidine in their primary structures (Figures 1,2). Thus, in these proteins the central motif of the classic NlpC/P60 proteins with the conserved glycine was the most amino-terminal conserved motif, followed by the motif containing the conserved histidine, and then the one with the cysteine. This arrangement suggested that there was a circular permutation with respect to the classical NlpC/P60 module in these proteins. All the NlpC/P60 homologs recovered in the above searches were collated and clustered at various thresholds of BLAST-bit-score to high-scoring pairs (HSP)-length ratios using the BLASTCLUST program. Multiple sequence alignments were prepared for the individual clusters using T-Coffee [35] and their secondary structure was predicted with the PHD program [36]. A comparison revealed that the core secondary structures of both the classical NlpC/P60 modules as well as those with permuted motifs showed 90% or greater concordance. These observations, taken together, suggest that classical NlpC/P60-like peptidases, the LRAT-like permuted forms, and the divergent GSPS amidase domain-like proteins comprise a vast superfamily of enzymes with different biochemical activities.

Figure 1 *(see figure on the next page)*
Multiple sequence alignment of the NlpC/P60 superfamily. Multiple sequence alignments of the different families of NlpC/P60 were constructed using T-Coffee [35] after parsing high-scoring pairs from PSI-BLAST search results. The PHD-secondary structure [36] is shown above the alignment with E representing a β strand, and H an α helix. The 85% consensus shown below the alignment was derived using the following amino acid classes: hydrophobic (h, ALICVMYFW, yellow shading), the aliphatic subset of the hydrophobic class are (l, ALIVMC, yellow shading), small (s, ACDGNPSTV, green) and polar (p, CDEHKNQRST, blue). A 'G', 'D', 'C' or 'N' shows the completely conserved amino acid in that group. The numbers colored light blue show the region where a zinc ribbon domain is inserted in the *Arabidopsis* LRAT homologs. The catalytic residues are highlighted in red, while the conserved cysteine present in LRAT and related proteins is colored red. Specific columns of residues that are peculiar to a particular category of NlpC/P60 (see text) are colored red. The consensus of the individual families and the entire superfamily are shown, and the subgroups of each family are separated by a space. The limits of the domains are indicated by the residue positions, in bold, on each side, or internally in the case of the permuted versions. A dotted line separates the amino and carboxyl termini of the permuted versions. The numbers within the alignment are non-conserved inserts that have not been shown. The sequences are denoted by their gene name followed by the species abbreviation and GeneBank identifier (gi). An alignment covering members of the bacterial P60 family is available in PFAM [78]. The species abbreviations are: Af, *Archaeoglobus fulgidus*; BPDP1, bacteriophage Dp-1; Ana, *Anabaena*; Atu, *Agrobacterium tumefaciens*; Ban, *Bacillus anthracis*; Bha, *Bacillus halodurans*; Bs, *Bacillus subtilis*; Bmel, *Brucella melitensis*; Cac, *Clostridium acetobutylicum*; Ccr, *Caulobacter crescentus*; Cj, *Campylobacter jejuni*; Ct, *Chlamydia trachomatis*; Cpn, *Chlamydophila pneumoniae*; Dr, *Deinococcus radiodurans*; Ec, *Escherichia coli*; Hi, *Haemophilus influenzae*; Hp, *Helicobacter pylori*; Lla, *Lactococcus lactis*; Lin, *Listeria innocua*; Lmo, *Listeria monocytogenes*; Mle, *Mycobacterium leprae*; Mlo, *Mesorhizobium loti*; Mtu, *Mycobacterium tuberculosis*; Nm, *Neisseria meningitidis*; Pae, *Pseudomonas aeruginosa*; Pmu, *Pasteurella multocida*; Rsol, *Ralstonia solanacearum*; Rrhi, *Rhizobium rhizogenes*; St, *Salmonella typhimurium*; Sme, *Sinorhizobium meliloti*; Sa, *Staphylococcus aureus*; Scoe, *Streptomyces coelicolor* A3; Spn, *Streptococcus pneumoniae*; Spy, *Streptococcus pyogenes*; Ssp, *Synechocystis* sp.; Tm, *Thermotoga maritima*; Vch, *Vibrio cholerae*; Xaxo, *Xanthomonas axonopodis*; Xf, *Xylella fastidiosa*; Ype, *Yersinia pestis*; At, *Arabidopsis thaliana*; Ce, *Caenorhabditis elegans*; Cfas, *Crithidia fasciculata*; Hs, *Homo sapiens*; Mm, *Mus musculus*; Rn, *Rattus norvegicus*; Tcr, *Trypanosoma cruzi*; CV, cowpox virus; FPV, fowlpox virus; MCV, molluscum contagiosum virus; MV, myxoma virus; RFV, rabbit fibroma virus; ShPV, sheeppox virus; SPV, swinepox virus; VV, vaccinia virus; VarV, variola virus; YDV, Yaba-like disease virus; AMV, *Amsacta moorei* entomopoxvirus; MSV, *Melanoplus sanguinipes* entomopoxvirus; AiV, Aichi virus; BCV, bovine calicivirus; ChV, Chiba virus; LV, Lordsdale virus; NorV, Norwalk virus; ShV, Southampton virus; Aev, avian encephalomyelitis virus.

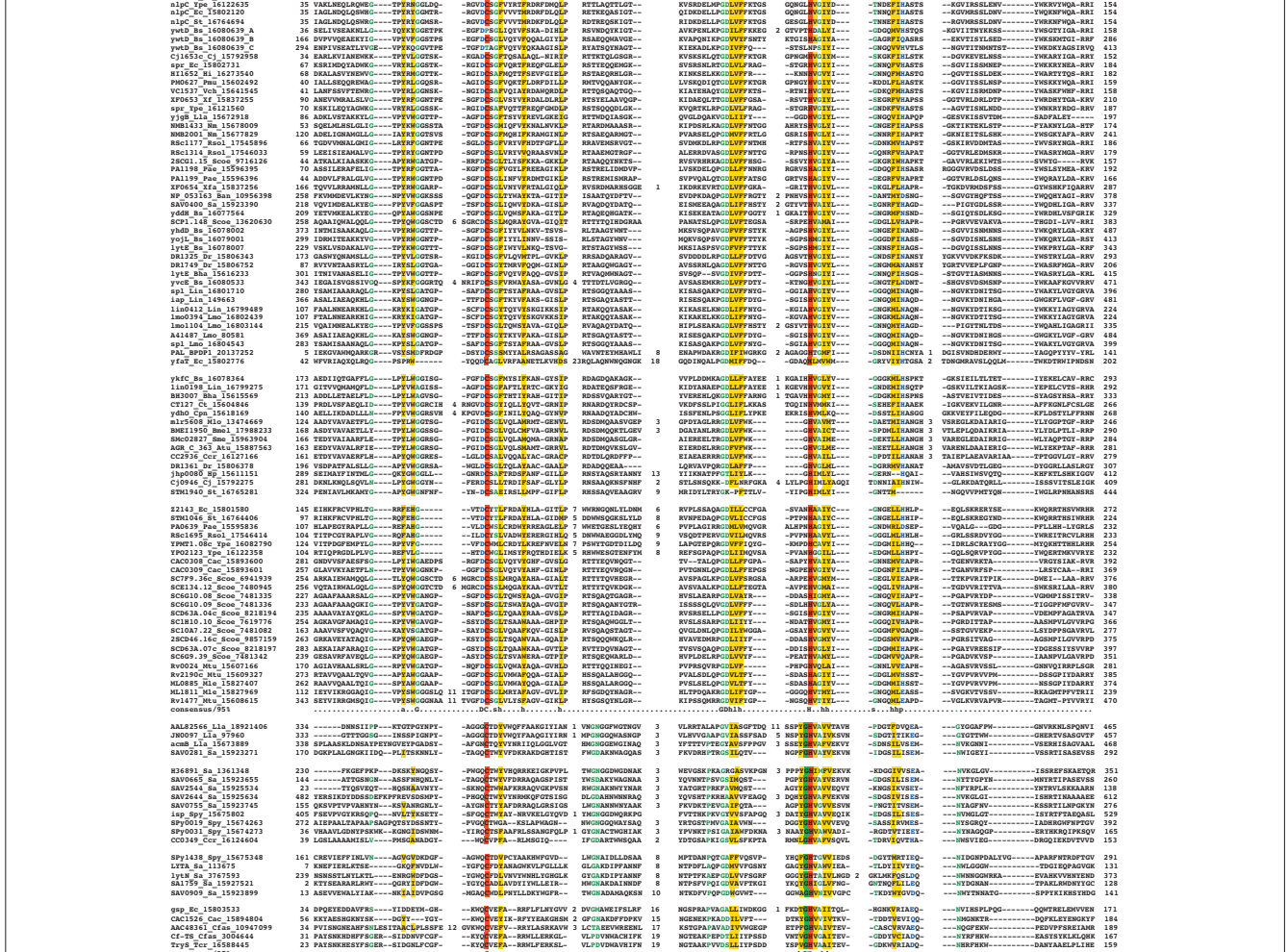

Figure 1 *(see legend on the previous page)*

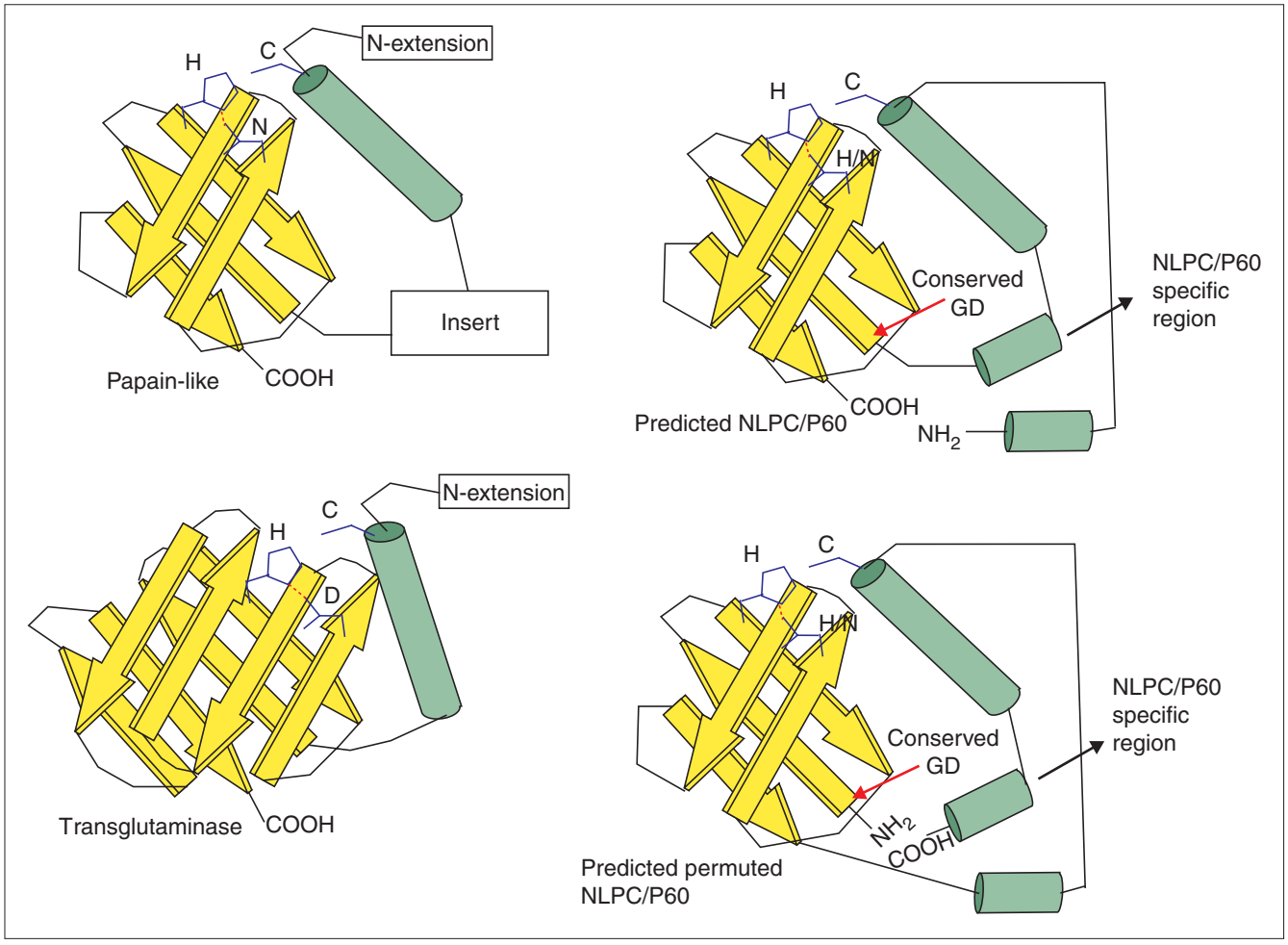

Figure 2
Topology diagram of the NlpC/P60 and structurally related proteases. The topology of the papain-like proteases (PDB:1ppn) and transglutaminases (PDB:1fie) are derived from the available crystal structures of these proteins. The predicted topology of the NlpC/P60 family members was derived from secondary structure and sequence conservation profile. Yellow arrows represent β strands and green cylinders represent α helices.

Structural and biochemical features of the NlpC/P60 superfamily

Secondary-structure prediction, based on multiple sequence alignments, revealed that the NlpC/P60 superfamily domains adopt an α+β fold with segregated α and β elements (Figure 1). The first recognizable secondary structure element in this domain is an α-helix that is present in practically all members of this superfamily, with the exception of a few members of the AcmB family from *Staphylococcus aureus* (Figures 1,2), where it appears to have entirely degenerated. The most prominent α-helix, which is conserved in all proteins of this superfamily, is the second helix, which is associated with the first conserved motif. The conserved cysteine occurs at the extreme amino terminus of this helix and is typically preceded by a polar residue. The central motif with the conserved glycine and the motif bearing the conserved histidine correspond to the two well conserved strands. Three more strands are predicted to occur carboxy-terminal to these two strands. Of these, the strand that immediately follows the strand with the conserved histidine, contains a conserved polar residue (either histidine or an acidic or amide residue). By analogy with all other thiol proteases, in the NlpC/P60 superfamily peptidases the cysteine would act as the nucleophile that attacks the peptide bond, while the histidine acts in sequential steps as the base and then acid catalyst for the proton transfers [37]. In other thiol proteases the histidine is oriented to a hydrogen bond with the cysteine by a third polar residue [37]. The presence of a third conserved polar position in the strand that follows the histidine-containing strand suggests that it serves as the orienting residue for the catalytic histidine in the NlpC/P60. While there is no directly detectable sequence similarity between the NlpC/P60 proteins and other superfamilies of peptidases, the arrangement of catalytic residues, with respect to the conserved secondary elements, in this superfamily is the same as that in several other peptidase superfamilies

(Figure 2). The configuration, comprising a catalytic cysteine at the amino terminus of a helix packed against a core three-stranded β-sheet, with the second and third strands bearing the catalytic histidine and its orienting polar partner, is seen in the transglutaminases and a related class of amino-group acetyltransferases, papain-like proteases, adenoviral proteases and ubiquitin carboxy-terminal hydrolases [32,38] (Figure 2). This suggests that despite the absence of significant sequence similarity, the NlpC/P60 domain is likely to contain a structural core corresponding to the classical papain-like protease fold (Figure 2).

In the case of papain-like proteases - calpains, transglutaminases and NH_2-acetyltransferases - the three-stranded sheet of the core is incorporated into β-barrel with five to eight constituent strands (Figure 2). Contrary to this, in the case of the ubiquitin carboxy-terminal hydrolases and the adenoviral proteases there is a tendency for the sheet to be incorporated into a scaffold with other α-helical elements forming a mixed α+β structure. The observation that the three-stranded unit of the NlpC/P60 domain is followed by two other β-strands, suggests that the NlpC/P60 proteins are likely to form a five-stranded barrel similar to that seen in papain-like proteases (Figures 1,2). However, in the NlpC/P60 domain, the distance between the helix bearing the catalytic cysteine and the core barrel is similar to that seen in the transglutaminases and NH_2-aminotransferases, suggesting that it lacks the large insert that is seen in several papain-like proteases (Figure 2). This difference may also explain the failure of sequence-structure threading algorithms, such as the combined fold prediction and 3DPSSM [39], to recover any significant hits to the papain-like proteases. Thus, the NlpC/P60 domain is structurally close to the minimal ancestral unit of the thiol-protease fold.

The prediction of a classical thiol-protease fold for the NlpC/P60 proteins also explains the circular permutation of the conserved motifs that is observed in several members of this superfamily such as LRAT and YaeF (Figure 1). Precedent for such a permutation is offered by the adenoviral proteases, in which the β-sheet with the catalytic histidine precedes the α-helix with the catalytic cysteine [40,41]. In this fold, the carboxyl terminus of the β-structure (barrel or sheet) with the histidine typically occurs in close proximity to the amino-terminal region of the domain (Figure 2). This could allow a permutation to survive without distorting the overall spatial arrangement of the catalytic residues. A comparison of the conserved residues and predicted secondary structure of the 'regular' NlpC/P60 domain with the permuted versions suggests that the five-stranded β-barrel is maintained intact, with the remaining segment including the helix with the cysteine being linked to either the amino or carboxyl terminus of the barrel (Figures 1,2).

The only biochemically characterized eukaryotic representative of this superfamily is LRAT, which transfers an acyl group from the *sn-1* position of phosphatidylcholine (lecithin) to retinal, to form retinyl esters [27,42,43]. The general base of the enzyme (cysteine) accepts the acyl group from lecithin, and this is followed by retinol being directed to attack the acyl-enzyme by the catalytic histidine followed by the release of the retinyl ester [42,43]. The similarity of this reaction mechanism to that catalyzed by the catalytic-triad cysteine and serine proteases has been previously proposed [27,44]. The phylogenetic relationship we demonstrate between the LRAT family of enzymes and thiol peptidases shows that this catalytic triad was derived only once in evolution and re-used for different mechanistically similar reactions such as acyl transfer, transacylation or peptide-bond hydrolysis. The presence of an analogous situation in the transglutaminases [30,45] and the related NH_2-acyltransferases [46] suggests that diverse acetyl/acyltransferase activities have been repeatedly derived in course of the evolution of the papain-like fold. The presence of duplicated versions of the NlpC/P60 domain in some proteins (Figures 1,3) suggests that the permutation could have potentially occurred through a duplicate intermediate in which the either the amino or carboxyl terminus of each of the copies degenerated, leaving behind a stable permuted unit.

In the thiol-protease fold, variable-sized inserts are often seen in the region between the strand containing the histidine and the strand bearing the orienting polar residue. This feature is even observed in the NlpC/P60 superfamily, including the insertion of an entire zinc ribbon domain [47] in the case of the plant homologs of LRAT. The orienting polar residue is either an acidic or a polar amide residue in most members of the thiol-peptidase fold [32,37]. In the majority of the members of the NlpC/P60 superfamily, this residue is histidine, which in the appropriate environment can act equivalently to an acidic orienting residue. In the NlpC/P60 domains of the AcmB-like divergent group and the circularly permuted YaeF-like and poxviral groups, this residue is an acidic residue, as in most other thiol peptidases (Figure 1). However, in the AcmB-like group it is shifted by two positions with respect to the rest of the superfamily (Figure 1), suggesting a specific elongation of the strand bearing this residue in this lineage.

Functional, architectural and phylogenetic diversity of the NlpC/P60 superfamily

We derived an evolutionary classification of the NlpC/P60 superfamily (Figure 4) using a combination of BLAST-score-based clustering, conventional phylogenetic analysis with neighbor-joining, least-squares [48,49] and maximum-likelihood methods [50], and parsimony analysis based on shared derived characters (synapomorphies). We initially performed single-linkage clustering of all the NlpC/P60 domains using the BLASTCLUST program with different thresholds of bit-score/length values. This allowed us to identify all the major, distinct lineages within the superfamily and also the individual ortholog groups that were conserved across various groups of organisms. From these preliminary clusters, we

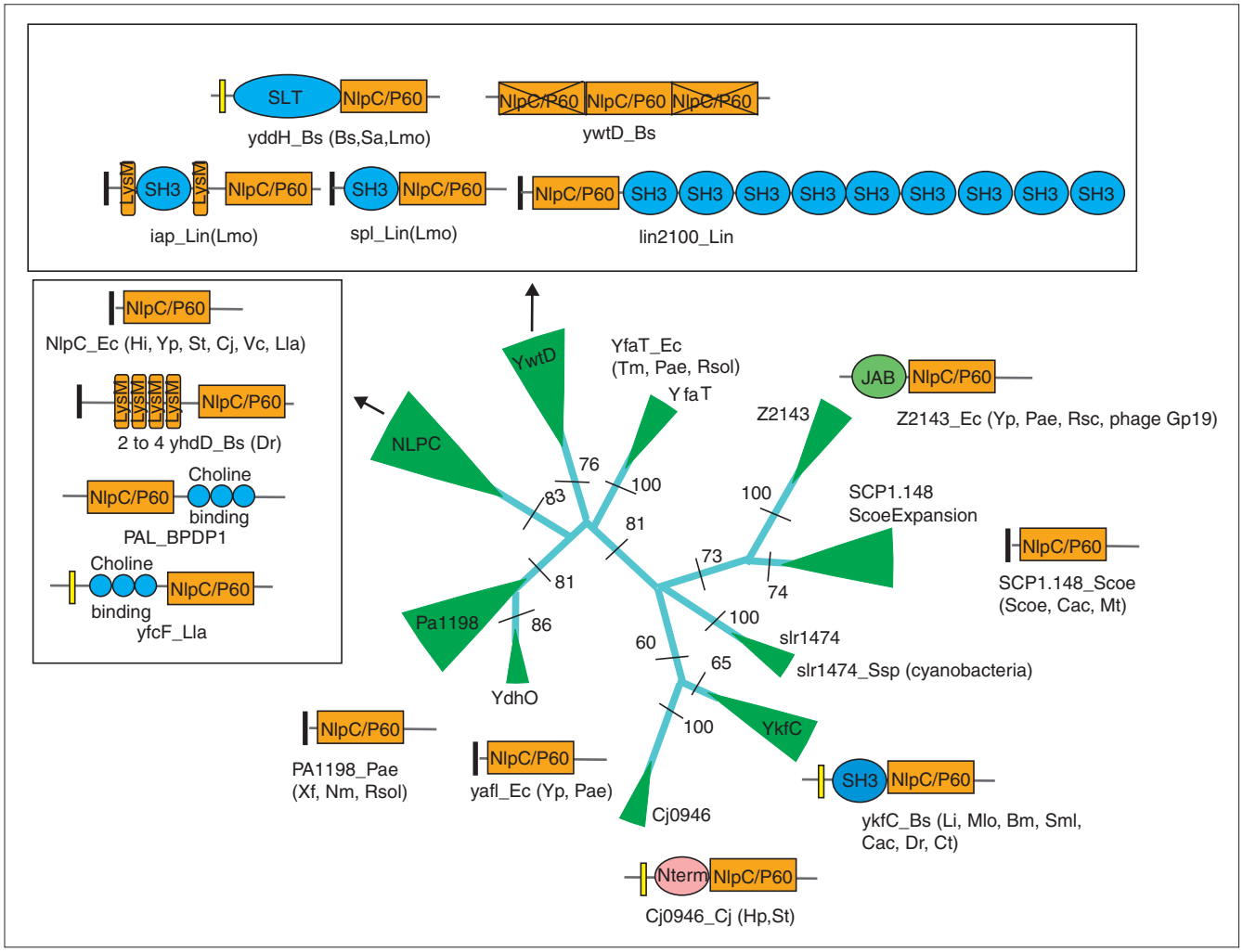

Figure 3
Phylogenetic relationships and domain architectures of NlpC/P60 family proper. The phyletic pattern of each family is shown next to the clade. The RELL bootstrap values for the major branches are shown at their base. The width of a given clade is approximately proportional to number proteins contained within it. N-term is a specific amino-terminal module that is restricted to the Cj0946 clade. The domain abbreviations are: SLT, soluble lytic transglycosidase. The signal peptides are shown as black rectangles, while the transmembrane regions are shown as yellow rectangles. The species abbreviations are given in the legend to Figure 1.

chose representatives and used them for rigorous phylogenetic analysis with the methods mentioned above. We also identified conserved sequence signatures and used them as characters to establish higher-order relationships.

Both the single-linkage clustering and the phylogenetic analysis identified four major families (Figure 4) within the NlpC/P60 superfamily: the P60-like family, the AcmB/LytN-like divergent family, the LRAT-like family of circularly permuted domains, and the YaeF/Poxvirus G6R type circularly permuted domains.

The P60-like family

The P60-like family typified by P60 and its obvious relatives includes the most commonly occurring versions of the superfamily, which are seen in most bacterial lineages (Figures 3,4). All members of this family have either a signal peptide or a transmembrane region, indicating an extracellular location (Figure 3). All characterized members of this family are peptidases, and they either hydrolyze the D-γ-glutamyl-meso-diaminopimelate linkage or *N*-acetylmuramate-L-alanine linkage [19,25]. There are 10 clearly identifiable orthologous lineages within this family which have varying phyletic distributions in bacteria (Figure 3). Some of these, like the YwtD-like orthologous group, are prevalent only in Gram-positive bacteria, while the NlpC-like orthologous group is present in both Gram-positive bacteria and proteobacteria. The Pal lineage of the NlpC-like orthologous group is restricted to the bacteriophages and their prophage remnants in Gram-positive bacteria. They

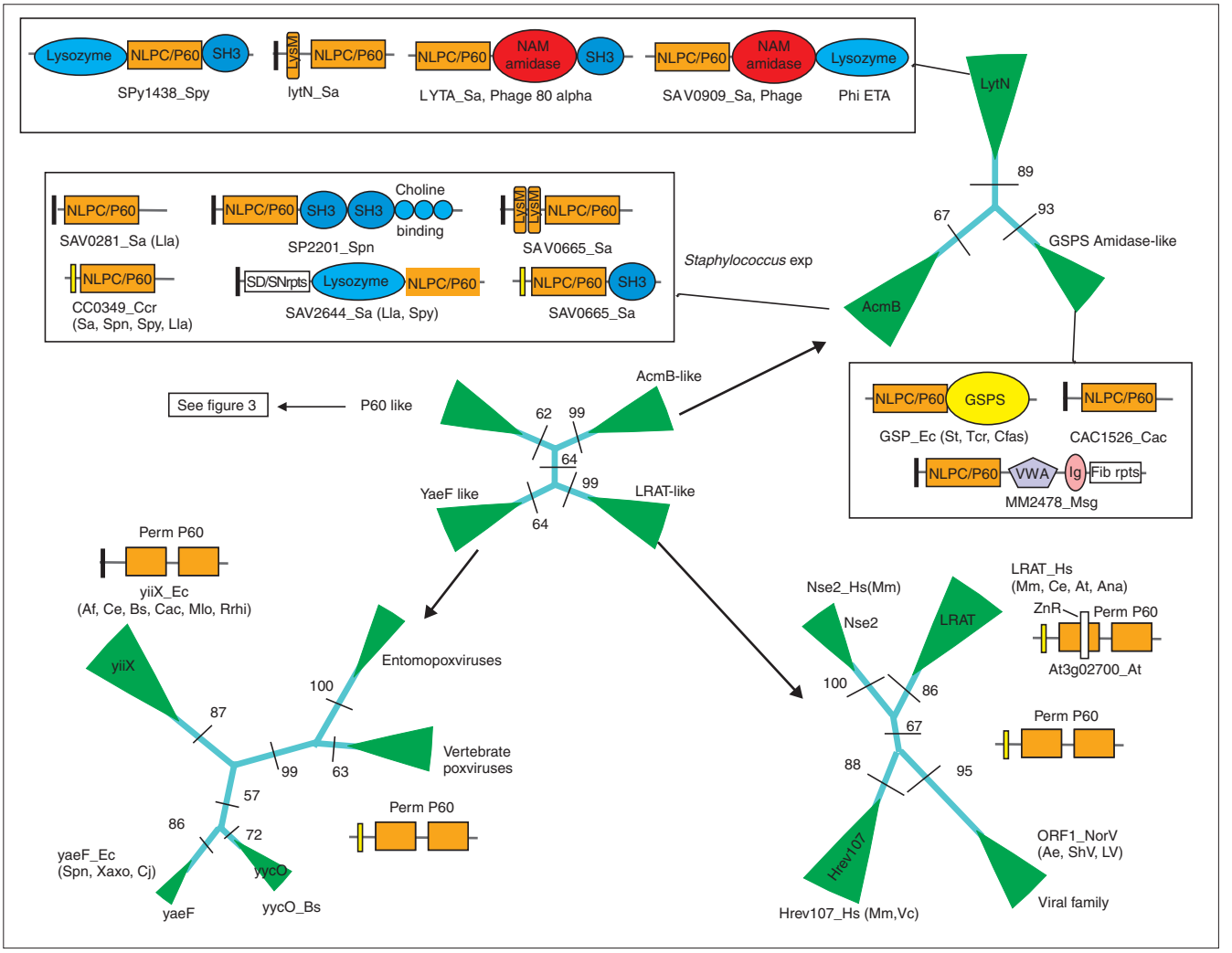

Figure 4
Phylogenetic relationships and domain architectures of NlpC/P60 superfamily. The domain abbreviations are: AM amidases, *N*-acetylmuramidase; VWA, von Willebrandt factor A; GSPS, glutathionylspermidine synthase; IG, immunoglobulin domain; Fib rpts, fibrinogen-like repeat domains; ZnR, inserted zinc ribbon. The conventions are the same as in Figure 3.

hydrolyze the *N*-acetylmuramate-L-alanine linkage, and are potent bacteriolytic enzymes [24,25]. They appear to represent a case where a host enzyme has been exapted (re-used) by a virus for penetrating the bacterial cell wall.

The presence of the P60-like family in most bacterial lineages suggests that a representative of this lineage was probably present before the divergence of the bacteria from a common ancestor. However, it appears to have differentiated to several distinct orthologous groups only later in bacterial evolution. Several orthologous groups show a patchy phyletic pattern: for example the orthologous group SCP1.148 is present in the actinomycetes and the Gram-positive bacterium *Clostridium acetobutylicum*. It shows a lineage-specific expansion in the former taxon, with 11 and 3 closely related, paralogous forms in the genomes of *Streptomyces* and *Mycobacterium tuberculosis*, respectively. Such phyletic patterns, together with the presence of members of this family in genomes of phages and prophages, suggests that lateral transfer was involved in the dissemination of some of these orthologous groups. Gene loss is an alternative possibility that could in part explain the sporadic distribution of some of these orthologous groups. If the trees showed a consistent clustering of proteins only from bacteria with large genomes, in which gene loss is less pervasive, then gene loss is likely to be a major factor. However, this pattern in not observed (Figures 3,4), suggesting an important role for lateral transfer.

In addition to the variable phyletic patterns, the proteins of the P60-like family also show extensive lineage-specific diversification in terms of their domain architectures

(Figure 3). In many proteins, the catalytic peptidase domain is fused to domains such as SH3, LysM, and choline-binding domains [51,52] (Figure 3). These domains probably aid them in interactions with peptides, carbohydrates and lipids that are associated with the bacterial cell wall. The NlpC/P60 domain is also fused to other catalytic domains such as the polysaccharide-hydrolyzing lysozyme domain, and the JAB domain that has recently been described to have metallopeptidase activity [53,54]. These are likely to function as two-headed enzymes that simultaneously attack different linkages in the murein. These architectures point to the potential diversity in the biological functions of the autolysins in bacterial cell-wall metabolism. They could be functional at different spatial locations and at different temporal points in the life cycle, thereby contributing to the diversity in the morphology of the bacterial cell and colonies.

The AcmB/LytN-like family

The AcmB/LytN-like family is the most divergent family of the NlpC/P60 superfamily (Figure 4). The monophyly of this family is strongly supported in the phylogenetic analysis (RELL bootstrap: 99%) and by synapomorphies, such as the position of the orienting polar residue (see above). Two major subfamilies within this family, typified respectively by AcmB and *S. aureus* LytN, are predominantly restricted to the Gram-positive bacteria, and function as cell-wall hydrolases. The AcmB subfamily shows a lineage-specific expansion in the genome of *S. aureus*, with around seven paralogs. Both the subfamilies exhibit considerable diversity of domain architectures, with most of the architectures mirroring those seen in the P60-like family. Thus, these proteins contain fusions of the NlpC/P60 peptidase domain with the SH3, choline-binding, LysM, lysozyme and NAM amidase domains (Figure 4), and this implicates them in different processes in cell-wall dynamics. A third subfamily of the AcmB/LytN-like family is typified by the amino-terminal amidase domain of the GSPS from γ-proteobacteria and trypanosomes [26,55,56]. GSPS is a bifunctional enzyme: its synthase domain carries out the ATP-dependent fusion of two glutathione moieties to the polyamine spermidine to form the antioxidant metabolite GSP, while its amidase domain hydrolyzes the amide linkage to release glutathione and spermidine [26,55,56]. The proteomes of *Methanosarcina* and the Gram-positive bacterium *C. acetobutylicum* have a member each of this GSPS-like subfamily. Both these proteins are secreted proteins, with the version from *Methanosarcina*, MM2478, additionally containing von Willebrandt factor A, immunoglobulin and fibrinogen-like repeat domains (Figure 4).

The phyletic distribution suggests that the AcmB/LytN-like family emerged early in the Gram-positive lineage, through rapid sequence divergence. With the exception of the GSPSs, all the proteins within each subfamily of this family are secreted or membrane-associated. Even the closest relatives of the GSPSs, namely MM2478 and CAC1526, are secreted proteins that are likely to function as extracellular peptidases. This suggests that even the GSPS subfamily probably arose within the Gram-positive bacteria followed by a lateral transfer to the archaeon *Methanosarcina* and the γ-proteobacteria. In the latter lineage, it was exapted to function as an intracellular enzyme, probably due to the similarity in structure of GSP and the ancestral extracellular substrate, namely the peptidoglycan peptides. MM2478 is a rare instance of an NlpC/P60 superfamily member occurring in the archaeal lineage. Its domain architecture suggests that it might function as a protease in the maturation of the complex extracellular matrix of this multicellular archaeon.

The YaeF/Poxvirus G6R–like and the LRAT-like families

Both the LRAT-like and the YaeF/Poxvirus G6R families are unified by the circular permutation that occurs at essentially the same point in both these families. The YaeF/Poxvirus G6R shows a peculiar phyletic distribution of being present only in bacteria, the archaeon *Archaeoglobus*, *C. elegans* and the poxviruses (Figure 4). The synapomorphy that unifies this family is the presence of an acidic orienting residue in the catalytic triad (Figure 1). Given the widespread distribution of this family in bacteria, with multiple distinct lineages, it is likely that this family arose early in the bacteria through a circular permutation from an ancestral, classical NlpC/P60 domain. Subsequently, it appears to have been transferred from a bacterial source to *Archaeoglobus*, the nematode lineage, and the common ancestor of all poxviruses. The high degree of conservation and maintenance of this protein across both entomopoxviruses and vertebrate poxviruses suggests that it performs an important function in these viruses (Figures 1,4). Along with the LRAT proteins, these proteins contain a second conserved cysteine (red letters in Figure 1), carboxy-terminal to the catalytic cysteine. This second cysteine has been shown to be critical for a functional LRAT enzyme, suggesting that the poxviral proteins may have a related activity (see below) [42]. Formation of acyl conjugates during the maturation of the lipoprotein component of the viral capsid could be one such potential function. Alternatively, these proteins could function as proteases in the maturation of the viral particle.

The LRAT family is thus far found only in eukaryotes and animal viruses, with the sole exception of a single member from the proteomes of *Vibrio cholerae* and *Anabaena*. The LRAT and its obvious orthologs are found, in addition to the vertebrates, in other animals and plants. In vertebrates the enzyme has an important role in the storage and mobilization of retinol (vitamin A) as esters in peripheral tissues and for generating an intermediate in the synthesis of the rhodopsin chromophore in the visual tissues [27]. The ortholog of the vertebrate LRAT, Egl-29 in *C. elegans* has been implicated in the development of the vulva [57]. The Egl-29 function is required in the cell neighboring the one in which it is expressed, suggesting that it may have a role in generating a regulatory signal [57]. This could raise the

interesting possibility of Egl-29 synthesizing an ester similar to the retinyl esters that might function as a secreted differentiation signal during vulval development. The presence of orthologs in plants suggests that even plants may process retinol or related compounds similar to those in animals.

The vertebrates possess two other distinct subfamilies that appear to have been derived from the LRAT family (Figure 4). One of these is typified by the candidate tumor suppressor protein, H-rev107, which inhibits growth of tumors induced by the oncoprotein H-Ras [58]. H-rev107 has been shown to exist in two cellular isoforms, one of which is associated with the cell membrane [59,60]. This suggests that H-rev107 could potentially act as an acyltransferase that might modify membrane components. A homolog of H-rev107 is detectable in the pathogenic bacterium *V. cholerae,* and phylogenetic analysis suggests that it has been acquired through lateral transfer from a vertebrate source. Further experimental analysis of this bacterial protein could be of considerable interest, as it could potentially have a role in interactions with its vertebrate host. The second vertebrate-specific subfamily is typified by the uncharacterized human protein NSE2, which could potentially function in a similar capacity to H-rev107. However, in these proteins the catalytic cysteine is replaced by a serine (Figure 1). Such substitutions have previously been observed in other papain-like peptidases with the serine participating in a functional active site [61,62]. Multiple paralogs of both H-rev107 and NSE2 are seen in vertebrates, suggesting some functional diversification of these subgroups.

The viral homologs of H-rev107 were observed in picornaviruses such as human parechoviruses, Aichi virus and avian encephalomyelitis virus [63]. We also detected related proteins in human caliciviruses such as the Norwalk virus and related viruses (Figures 1,4). These proteins could potentially function as a second protease in the processing of viral polyproteins, or they could function as a viral enzyme that could modify some membrane component, like their cellular homologs. All the members of the LRAT family, including the forms from RNA viruses, clearly form a distinct family which excludes the YaeF/poxvirus G6R family, which shows a similar circular permutation. The widespread distribution of the LRAT family in eukaryotes suggests that they possessed the enzymes from an early stage in their evolution. The NlpC/P60 superfamily is largely absent from most archaeal proteomes. However, it is very widespread in bacteria, and includes forms with circular permutations, similar to the eukaryotic LRAT-like proteins. Hence, the LRAT family was probably acquired by the eukaryotes, by the lateral transfer of a bacterial precursor similar to the YaeF/YiiX family. This could have either occurred during the primary γ-proteobacterial endosymbiosis, which gave rise to the mitochondrion, or as a result of a subsequent transfer from some other bacterial source. The currently available sequences do not allow us to distinguish between these possibilities. The drastic change in biochemical function might have resulted in an accelerated evolution of these eukaryotic proteins, resulting in considerable divergence from the bacterial YaeF/YiiX-like forms.

Conclusions

We present a detailed evolutionary and structural analysis of the NlpC/P60 superfamily of peptidases. We show that this superfamily, in addition to the well characterized P60-related peptidoglycan peptidases, includes divergent cell-wall hydrolases typified by AcmB/LytN, eukaryotic proteins such as LRAT, Egl-26 and their relatives, amidase domains of GSPSs, and several other uncharacterized bacterial proteins. We identified three residues involved in catalysis, namely a cysteine, a histidine and a polar residue, that are conserved throughout this superfamily. A comparison of the predicted secondary structure for this superfamily with known peptidase folds revealed that these three residues show exactly the same arrangement as the catalytic triad of papain-like thiol peptidases. The NlpC/P60 domain is predicted to adopt a papain-like fold, in which the catalytic cysteine is at the amino terminus of a helix, which is packed against a five-stranded β-barrel with the histidine and the polar residue. The NlpC/P60 superfamily has diversified into four major families, with two of them - namely the LRAT and YaeF/Poxvirus G6R families - showing a circular permutation of the catalytic domain.

This analysis could considerably aid further experimental investigations of the NlpC/P60 superfamily. Certain members of this superfamily such as Pal have recently been proposed as potent antibacterial agents [24]. Identification of novel members of this superfamily in diverse bacterial proteomes could point to additional candidates with similar bacteriolytic activity. Furthermore, this analysis could help in identifying enzymes with important roles in bacterial colony and cell morphology. Recognition of Egl-26, H-rev107 and NSE2 as homologs of LRAT suggests a potential role for these proteins as novel components of lipid metabolism in eukaryotes. Further studies on the homologs of these proteins in diverse animal viruses and the pathogenic bacterium *V. cholerae* could help in uncovering hitherto unknown mechanisms by which these pathogens could interact with their hosts.

Materials and methods

The non-redundant (nr) database of protein sequences (National Center for Biotechnology Information (NCBI)) was searched using the BLASTP program [28]. Profile searches were conducted using the PSI-BLAST program with either a single sequence or an alignment used as the query, with a profile inclusion expectation (E) value threshold of 0.01, and was iterated until convergence [28,64]. Before use in PSI-BLAST searches, the NlpC domain was evaluated for

compositional bias using the SEG program [65]. No such bias that could skew the statistics of sequence relationships in searches of the nr database was detected. Accordingly, to achieve maximum sensitivity all searches were run with the compositional-bias-based statistics turned off [66]. Multiple sequence alignments were constructed using the T_Coffee program [35], followed by manual correction based on the PSI-BLAST results. Conserved motifs were searched for using the Gibbs sampling procedure. Homologs recovered in the searches were collated and clustered at various thresholds of BLAST-bit-score to HSP-length ratios using the BLASTCLUST program [67].

Structural manipulations were carried out using the Swiss-PDB viewer program [68] and the ribbon diagrams were constructed with MOLSCRIPT [69]. Searches of the PDB database with query structures was conducted using the DALI program [70,71]. Protein secondary structure was predicted using a multiple sequence alignment as the input for the PHD program [36,72]. Signal peptides were predicted using the SIGNALP program [73-75] and the transmembrane regions were predicted using the TOPRED program [76,77].

Phylogenetic analysis was carried out using the maximum-likelihood, neighbor-joining and least-squares methods [49,50]. Briefly, this process involved the construction of a least-squares tree using the FITCH program or a neighbor-joining tree using the NEIGHBOR program (both from the Phylip package) [48], followed by local rearrangement using the Protml program of the Molphy package [50] to arrive at the maximum-likelihood (ML) tree. The statistical significance of various nodes of this ML tree was assessed using the relative estimate of logarithmic likelihood bootstrap (Protml RELL-BP) with 10,000 replicates.

Additional data files

A copy of the alignment shown in figure 1 is available as an Additional data file in MS-WORD format with the online version of this article.

References

1. Moat AG, Foster JW (Eds): *Microbial Physiology*, 4th edn. New York: John Wiley; 2002.
2. Madigan MM, Parker J, Madigan MT, Martinko JM: *Brock Biology of Microorganisms*, 10th edn. Upper Saddle River, NJ: Prentice Hall; 2002.
3. van Heijenoort J: **Formation of the glycan chains in the synthesis of bacterial peptidoglycan.** *Glycobiology* 2001, **11:**25R-36R.
4. Healy VL, Lessard IA, Roper DI, Knox JR, Walsh CT: **Vancomycin resistance in enterococci: reprogramming of the D-Ala-D-Ala ligases in bacterial peptidoglycan biosynthesis.** *Chem Biol* 2000, **7:**R109-R119.
5. Born TL, Blanchard JS: **Structure/function studies on enzymes in the diaminopimelate pathway of bacterial cell wall biosynthesis.** *Curr Opin Chem Biol* 1999, **3:**607-613.
6. Atrih A, Foster SJ: **The role of peptidoglycan structure and structural dynamics during endospore dormancy and germination.** *Antonie Van Leeuwenhoek* 1999, **75:**299-307.
7. Bramhill D: **Bacterial cell division.** *Annu Rev Cell Dev Biol* 1997, **13:**395-424.
8. Smith TJ, Blackman SA, Foster SJ: **Autolysins of *Bacillus subtilis*: multiple enzymes with multiple functions.** *Microbiology* 2000, **146:**249-262.
9. Longchamp PF, Mauel C, Karamata D: **Lytic enzymes associated with defective prophages of *Bacillus subtilis*: sequencing and characterization of the region comprising the *N*-acetylmuramoyl-L-alanine amidase gene of prophage PBSX.** *Microbiology* 1994, **140:**1855-1867.
10. Schuch R, Nelson D, Fischetti VA: **A bacteriolytic agent that detects and kills *Bacillus anthracis*.** *Nature* 2002, **418:**884-849.
11. Foster SJ: **Analysis of *Bacillus subtilis* 168 prophage-associated lytic enzymes, identification and characterization of CWLA-related prophage proteins.** *J Gen Microbiol* 1993, **139:**3177-3184.
12. Walsh CT, Fisher SL, Park IS, Prahalad M, Wu Z: **Bacterial resistance to vancomycin: five genes and one missing hydrogen bond tell the story.** *Chem Biol* 1996, **3:**21-28.
13. Lessard IA, Pratt SD, McCafferty DG, Bussiere DE, Hutchins C, Wanner BL, Katz L, Walsh CT: **Homologs of the vancomycin resistance D-Ala-D-Ala dipeptidase VanX in *Streptomyces toyocaensis*, *Escherichia coli* and *Synechocystis*: attributes of catalytic efficiency, stereoselectivity and regulation with implications for function.** *Chem Biol* 1998, **5:**489-504.
14. McCafferty DG, Lessard IA, Walsh CT: **Mutational analysis of potential zinc-binding residues in the active site of the enterococcal D-Ala-D-Ala dipeptidase VanX.** *Biochemistry* 1997, **36:**10498-10505.
15. Murzin AG: **Structural classification of proteins: new superfamilies.** *Curr Opin Struct Biol* 1996, **6:**386-394.
16. Aravind L, Mazumder R, Vasudevan S, Koonin EV: **Trends in protein evolution inferred from sequence and structure analysis.** *Curr Opin Struct Biol* 2002, **12:**392-399.
17. Bussiere DE, Pratt SD, Katz L, Severin JM, Holzman T, Park CH: **The structure of VanX reveals a novel amino-dipeptidase involved in mediating transposon-based vancomycin resistance.** *Mol Cell* 1998, **2:**75-84.
18. Margot P, Wahlen M, Gholamhoseinian A, Piggot P, Karamata D, Gholamhuseinian A: **The lytE gene of *Bacillus subtilis* 168 encodes a cell wall hydrolase.** *J Bacteriol* 1998, **180:**749-752.
19. Ohnishi R, Ishikawa S, Sekiguchi J: **Peptidoglycan hydrolase LytF plays a role in cell separation with CwlF during vegetative growth of *Bacillus subtilis*.** *J Bacteriol* 1999, **181:**3178-3184.
20. Schubert K, Bichlmaier AM, Mager E, Wolff K, Ruhland G, Fiedler F: **P45, an extracellular 45 kDa protein of *Listeria monocytogenes* with similarity to protein p60 and exhibiting peptidoglycan lytic activity.** *Arch Microbiol* 2000, **173:**21-28.
21. Reinscheid DJ, Gottschalk B, Schubert A, Eikmanns BJ, Chhatwal GS: **Identification and molecular analysis of PcsB, a protein required for cell wall separation of group B streptococcus.** *J Bacteriol* 2001, **183:**1175-1183.
22. Ishikawa S, Hara Y, Ohnishi R, Sekiguchi J: **Regulation of a new cell wall hydrolase gene, cwlF, which affects cell separation in *Bacillus subtilis*.** *J Bacteriol* 1998, **180:**2549-2555.
23. Wuenscher MD, Kohler S, Bubert A, Gerike U, Goebel W: **The *iap* gene of *Listeria monocytogenes* is essential for cell viability, and its gene product, p60, has bacteriolytic activity.** *J Bacteriol* 1993, **175:**3491-3501.
24. Loeffler JM, Nelson D, Fischetti VA: **Rapid killing of *Streptococcus pneumoniae* with a bacteriophage cell wall hydrolase.** *Science* 2001, **294:**2170-2172.
25. Sheehan MM, Garcia JL, Lopez R, Garcia P: **The lytic enzyme of the pneumococcal phage Dp-1: a chimeric lysin of intergeneric origin.** *Mol Microbiol* 1997, **25:**717-725.
26. Oza SL, Ariyanayagam MR, Fairlamb AH: **Characterization of recombinant glutathioninylspermidine synthetase/amidase from *Crithidia fasciculata*.** *Biochem J* 2002, **364:**679-686.
27. Rando RR: **Membrane-bound lecithin-retinol acyltransferase.** *Biochem Biophys Res Commun* 2002, **292:**1243-1250.
28. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25:**3389-3402.
29. Aravind L, Koonin EV: **Classification of the caspase-hemoglobinase fold: detection of new families and implications for the origin of the eukaryotic separins.** *Proteins* 2002, **46:**355-367.

30. Anantharaman V, Koonin EV, Aravind L: **Peptide-N-glycanases and DNA repair proteins, Xp-C/Rad4, are, respectively, active and inactivated enzymes sharing a common transglutaminase fold.** *Hum Mol Genet* 2001, **10:**1627-1630.
31. Babe LM, Craik CS: **Viral proteases: evolution of diverse structural motifs to optimize function.** *Cell* 1997, **91:**427-430.
32. Barrett AJ, Rawlings ND: **Evolutionary lines of cysteine peptidases.** *Biol Chem* 2001, **382:**727-733.
33. Wang X, Mani N, Pattee PA, Wilkinson BJ, Jayaswal RK: **Analysis of a peptidoglycan hydrolase gene from *Staphylococcus aureus* NCTC 8325.** *J Bacteriol* 1992, **174:**6303-6306.
34. Sugai M, Fujiwara T, Komatsuzawa H, Suginaka H: **Identification and molecular characterization of a gene homologous to epr (endopeptidase resistance gene) in *Staphylococcus aureus*.** *Gene* 1998, **224:**67-75.
35. Notredame C, Higgins DG, Heringa J: **T-Coffee: A novel method for fast and accurate multiple sequence alignment.** *J Mol Biol* 2000, **302:**205-217.
36. Rost B, Sander C: **Prediction of protein secondary structure at better than 70% accuracy.** *J Mol Biol* 1993, **232:**584-599.
37. Walsh C: *Enzymatic Reaction Mechanisms*. New York: WH Freeman; 1995.
38. **Structural Classification of Proteins (SCOP) database** [http://scop.mrc-lmb.cam.ac.uk/scop]
39. Bujnicki JM, Elofsson A, Fischer D, Rychlewski L: **LiveBench-1: continuous benchmarking of protein structure prediction servers.** *Protein Sci* 2001, **10:**352-361.
40. Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L, Mitchell W, Olinger L, Tatusov RL, Zhao Q, *et al.*: **Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*.** *Science* 1998, **282:**754-759.
41. Mossessova E, Lima CD: **Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast.** *Mol Cell* 2000, **5:**865-876.
42. Mondal MS, Ruiz A, Bok D, Rando RR: **Lecithin retinol acyltransferase contains cysteine residues essential for catalysis.** *Biochemistry* 2000, **39:**5215-5220.
43. Mondal MS, Ruiz A, Hu J, Bok D, Rando RR: **Two histidine residues are essential for catalysis by lecithin retinol acyl transferase.** *FEBS Lett* 2001, **489:**14-18.
44. Jauhiainen M, Stevenson KJ, Dolphin PJ: **Human plasma lecithin-cholesterol acyltransferase. The vicinal nature of cysteine 31 and cysteine 184 in the catalytic site.** *J Biol Chem* 1988, **263:**6525-6533.
45. Makarova KS, Aravind L, Koonin EV: **A superfamily of archaeal, bacterial, and eukaryotic proteins homologous to animal transglutaminases.** *Protein Sci* 1999, **8:**1714-1719.
46. Sinclair JC, Sandy J, Delgoda R, Sim E, Noble ME: **Structure of arylamine N-acetyltransferase reveals a catalytic triad.** *Nat Struct Biol* 2000, **7:**560-564.
47. Aravind L, Koonin EV: **DNA-binding proteins and evolution of transcription regulation in the archaea.** *Nucleic Acids Res* 1999, **27:**4658-4670.
48. Felsenstein J: **PHYLIP - Phylogeny Inference Package (Version 3.2).** *Cladistics* 1989, **5:**164-166.
49. Felsenstein J: **Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods.** *Methods Enzymol* 1996, **266:**418-427.
50. Hasegawa M, Kishino H, Saitou N: **On the maximum likelihood method in molecular phylogenetics.** *J Mol Evol* 1991, **32:**443-445.
51. Ponting CP, Aravind L, Schultz J, Bork P, Koonin EV: **Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer.** *J Mol Biol* 1999, **289:**729-745.
52. Fernandez-Tornero C, Garcia E, Lopez R, Gimenez-Gallego G, Romero A: **Two new crystal forms of the choline-binding domain of the major pneumococcal autolysin: insights into the dynamics of the active homodimer.** *J Mol Biol* 2002, **321:**163-173.
53. Verma R, Aravind L, Oania R, McDonald WH, Yates IJ, Koonin EV, Deshaies RJ: **Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome.** *Science* 2002, **298:**611-615.
54. Cope GA, Suh GS, Aravind L, Schwarz SE, Zipursky SL, Koonin EV, Deshaies RJ: **Role of predicted metalloprotease motif of Jab1/Csn5 in cleavage of NEDD8 from CUL1.** *Science* 2002, **298:**608-611.
55. Bollinger JM Jr, Kwon DS, Huisman GW, Kolter R, Walsh CT: **Glutathionylspermidine metabolism in *Escherichia coli*. Purification, cloning, overproduction, and characterization of a bifunctional glutathionylspermidine synthetase/amidase.** *J Biol Chem* 1995, **270:**14031-14041.
56. Tetaud E, Manai F, Barrett MP, Nadeau K, Walsh CT, Fairlamb AH: **Cloning and characterization of the two enzymes responsible for trypanothione biosynthesis in *Crithidia fasciculata*.** *J Biol Chem* 1998, **273:**19383-19390.
57. Newman AP, Inoue T, Wang M, Sternberg PW: **The *Caenorhabditis elegans* heterochronic gene *lin-29* coordinates the vulval-uterine-epidermal connections.** *Curr Biol* 2000, **10:**1479-1488.
58. Sers C, Emmenegger U, Husmann K, Bucher K, Andres AC, Schafer R: **Growth-inhibitory activity and downregulation of the class II tumor-suppressor gene H-rev107 in tumor cell lines and experimental tumors.** *J Cell Biol* 1997, **136:**935-944.
59. Hajnal A, Klemenz R, Schafer R: **Subtraction cloning of H-rev107, a gene specifically expressed in H-ras resistant fibroblasts.** *Oncogene* 1994, **9:**479-490.
60. Akiyama H, Hiraki Y, Noda M, Shigeno C, Ito H, Nakamura T: **Molecular cloning and biological activity of a novel Ha-Ras suppressor gene predominantly expressed in skeletal muscle, heart, brain, and bone marrow by differential display using clonal mouse EC cells, ATDC5.** *J Biol Chem* 1999, **274:**32192-32197.
61. Gorbalenya AE, Donchenko AP, Blinov VM, Koonin EV: **Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. A distinct protein superfamily with a common structural fold.** *FEBS Lett* 1989, **243:**103-114.
62. Krem MM, Di Cera E: **Molecular markers of serine protease evolution.** *EMBO J* 2001, **20:**3036-3045.
63. Hughes PJ, Stanway G: **The 2A proteins of three diverse picornaviruses are related to each other and to the H-rev107 family of proteins involved in the control of cell proliferation.** *J Gen Virol* 2000, **81:**201-207.
64. Aravind L, Koonin EV: **Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches.** *J Mol Biol* 1999, **287:**1023-1040.
65. Wootton JC: **Non-globular domains in protein sequences: automated segmentation using complexity measures.** *Comput Chem* 1994, **18:**269-285.
66. Schaffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, Koonin EV, Altschul SF: **Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements.** *Nucleic Acids Res* 2001, **29:**2994-3005.
67. **BLASTCLUST** [ftp://ftp.ncbi.nih.gov/blast/documents/README.bcl]
68. Guex N, Peitsch MC: **SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling.** *Electrophoresis* 1997, **18:**2714-2723.
69. Kraulis PJ: **Molscript.** *J Appl Crystallogr* 1991, **24:**946-950.
70. Holm L, Sander C: **Protein structure comparison by alignment of distance matrices.** *J Mol Biol* 1993, **233:**123-138.
71. **EMBL DALI** [http://www.ebi.ac.uk/dali]
72. **PHD** [http://cubic.bioc.columbia.edu/predictprotein/submit_adv.html]
73. Nielsen H, Engelbrecht J, Brunak S, von Heijne G: **A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.** *Int J Neural Syst* 1997, **8:**581-599.
74. Nielsen H, Engelbrecht J, Brunak S, von Heijne G: **Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.** *Protein Eng* 1997, **10:**1-6.
75. **SignalP** [http://www.cbs.dtu.dk/services/SignalP-2.0]
76. von Heijne G: **Membrane protein structure prediction: hydrophobicity analysis and the 'positive inside' rule.** *J Mol Biol* 1992, **225:**487-494.
77. **TopRed** [http://bioweb.pasteur.fr/seqanal/interfaces/toppred.html]
78. **PFAM** [http://www.sanger.ac.uk/Software/Pfam/index.shtml]