number: 2635222), several proteins from Clostridium perfringens (including a hyaluronidase), and a putative serine/threonine kinase from Synecocystis sp. Many of the bacterial proteins identified are from intracellular pathogens that infect eukaryotic cells and probably are involved in cell invasion.

Threading calculations and model building provide convincing evidence that the N-terminus of the P60 invasion protein has an SH3 fold. The UCLA fold-recognition server predicted that P60_LISGR contains a region that has a fold similar to that of the SH3 domain of the putative proteoglycan tyrosine kinase [PDB entry: 1shs (Z = 6.70, which is well above the confidence threshold of 5.0 \leq Z \leq 5)]. In addition, eight out of the ten highest-scoring results had folds homologous to SH3 domains; the two highest-scoring – both of this domains – had Z scores of \( Z = 5.0 \). A second fold-recognition server, THREADER2 (Ref. 5), returned as the two highest-scoring results 1shs (the SH3 domain from the FYN proto-oncogene tyrosine kinase; \( Z = 7.68 \)) and 1shs (the SH3 domain from \( \alpha \)-spectrin; \( Z = 6.81 \)). Both scores are well above the ‘very significant’ threshold for THREADER2 (\( Z = 3.5 \)). The next-best result, 1mjc (the major cold shock protein 7.4 of Escherichia coli), which does not contain an SH3 domain, had a substantially lower score (\( Z = 3.0 \)).

We built a model of the fragment of P60_LISGR based on the chicken SRC tyrosine kinase\(^{133} \), using the alignment shown in Fig. 1. All residues boxed in the chicken SH3 domain structure correspond to hydrophobic residues (or threonine or glycine residues) in P60_LISGR. An asparagine residue that replaces the conserved proline residue present in the eukaryotic SH3 domains (shown in Fig. 1) is exposed and lies at the bottom of the groove in SH3 domains that bind peptides. The GTPase-activating protein GTPA RAT and other SH3 homologues have a valine residue at this position, which shows that the proline residue is not essential.

**Functional significance.** Invasion of eukaryotic cells by most pathogenic bacteria is accompanied by tyrosine phosphorylation, and inhibition of tyrosine phosphorylation impairs invasion by Listeria monocytogenes\(^{134} \). Listeria contains several invasion proteins. Different invasion factors – sometimes in concert – facilitate invasion of different cell types. P60 is important for invasion of epithelial cells\(^{135} \) and also for survival within the host cell. Indeed, the N-terminus of members of the P60 family of invasion proteins are highly conserved among different species of Listeria, which implies that this region is functionally important.

The P60 protein itself is thought to be a murine hydrolase\(^{136} \). It consists of three domains: the conserved N-terminus, which we suggest is an SH3 domain; a central domain that contains Ser/Thr-rich repeats; and a C-terminal domain, which is homologous to a number of a-amylases and starch-degrading enzymes. Species of bacteria that contain homologues of the putative SH3 domain from P60_LISGR are pathogens that invade eukaryotic cells. The SH3 domains of these prokaryotes might therefore have two possible functions: (1) promoting survival of a pathogen within the invaded cell by modulating pathways controlled by SH3 domains; or (2) promoting invasion by binding to receptors on eukaryotic cells.

**Conclusions.** We have suggested, on the basis of sequence similarity, structural compatibility and function, that P60_LISGR contains an SH3 domain. If this is confirmed, the appearance of SH3 domains in L. grayi will extend the range of this important family of proteins to prokaryotes (see Box 1).

---

### References

2 Patone, C. J. et al. (1994) Biochemistry 33, 6052–6062
10 Waisscher, M. D. et al. (1999) J. Bacteriol. 175, 3496–3504

### Acknowledgements

We thank C. Chothia, S. Teichmann and R. Williams for suggestions, and the Wellcome Trust and BBSRC (grant 977144) for support.

**Box 1. Note added in proof**

After this manuscript was submitted, Blilwe et al. reported the structure of SH3-like domains in CheA, a histidine kinase from the bacterium Thermotoga maritima that is homologous to proteins from Escherichia coli. The two SH3-like regions in CheA are ‘domain swapped’, which alters the sequence pattern.

Blilwe et al. (1999) Cell 96, 141–199
upstream of this ORF. Note that a construct that contains only 160 upstream nucleotides does not provide resistance to the above-mentioned amino acids. The upstream sequence does not contain a stop codon in frame with ORF f138. Moreover, one of the ATG codons in this sequence is preceded by a predicted ribosome-binding site. We designated the resultant, extended ORF (62160–61546 bp in M87049) \( rhtB \).

Disruption of the chromosomal \( rhtB \) gene causes hypersusceptibility to homoserine lactone and homoserine (V. V. Aleshin, unpublished). The RhtB protein is predicted to be highly hydrophobic and to possess six transmembrane segments. We have found a set of proteins that are homologous to RhtB in a wide range of prokaryotes that includes proteobacteria, cyanobacteria, bacilli and mycobacteria, and the archaea Archaeoglobus fulgidus and Methanobacterium thermoautotrophicum (Fig. 1). We performed a PSI-BLAST search of the non-redundant database at the NCBI and gapped BLAST searches of unfinished microbial genomes. A PSI-BLAST search, with an \( E \)-value threshold of \( 10^{-3} \), retrieved a set of proteins in three iterations – after which the search converged. In a gapped BLAST search, the probabilities of chance matches were estimated for the most-closely related sequences (\( p < 10^{-15} \)) and the most-distantly related (\( p < 10^{-3} \)) sequences. Most of the sequences homologous to the RhtB sequence represent hypothetical transmembrane proteins, some of which recently have been included in the UPF0048 family. We suggest that RhtB is involved in the efflux of homoserine and threonine in \( E. coli \).

We generated unrooted dendrograms by neighbour-joining and maximum-parsimony methods, using the PHYLIP 3.572 package with bootstrap analysis. Dendrograms (not shown) indicate that two different subfamilies exist: an RhtB-related subfamily and a LysE-related subfamily (Fig. 1).

### Figure 1

Multiple alignment of RhtB proteins. The fragments listed were selected from >60 sequences on the basis of the maximum dissimilarity in their primary structures. The distances between the motifs and the distances from the protein termini are indicated. Where >50% of sequences have similar or identical residues at a given position, a consensus residue is assigned (a, aromatic residue (F, Y or W); b, bulky aliphatic residue (L, I, V or M); b, bulky aliphatic/aromatic residue (L, V, M, F, Y or W); s, small residue (G, S, T or A); c, positively charged residue (K, R or H). Conserved residues are highlighted in colour: red indicates residues that fit the general consensus well; yellow indicates residues that fit the general consensus to a lesser extent; blue indicates residues that fit the RhtB subfamily consensus; green indicates residues that fit the LysE subfamily consensus. The positions of predicted transmembrane helices are shown as thick black lines.

Accession numbers in databases (gb, GenBank; gi, gene identification; sp, SWISS-PROT) or the contributing genome centers for sequences of unfinished genomes (GTC, Genome Therapeutics Corporation; OUACGT, University of Oklahoma Advanced Center for Genome Technology; Sanger, Sanger Centre; TIGR, The Institute for Genomic Research) are indicated in the right-hand column. Feature tables of the items shown in brackets were modified by either shifting the translation-initiation point or partial alteration of the reading frame. Aa, Actinobacillus actinomycetemcomitans; Al, Archeoglobus fulgidus; Ba, Bacillus subtilis; Bo, Bordetella pertussis; Bu, Bacillus subtilis; Ca, Clostridium acetobutylicum; Cg, Corynebacterium glutamicum; Cp, Campylobacter jejuni; Cj, Chlamydia jejuni; Dr, Deinococcus radiodurans; Ec, Escherichia coli; Hi, Haemophilus influenzae; Hp, Helicobacter pylori; Mt, Methanobacterium thermoautotrophicum; My, Mycobacterium tuberculosis; Pa, Pseudomonas aeruginosa; Pp, Pseudomonas putida; Ps, Pseudomonas syringae; Rc, Rhodobacter capsulatus; Sc, Shewanella oneidensis; Sf, Shewanella colombiana; Sy, Synechocystis sp. PCC 6803; Tm, Thermotoga maritima; Vc, Vibrio cholerae; Vp, Yersinia pestis.
genomes encode several paralogs from the two subfamilies (e.g. Bacillus subtilis, E. coli and Pseudomonas aeruginosa encode three, six and 12 paralogs, respectively). Thus, the divergence between the subfamilies is associated with gene duplication rather than with taxonomic diversification and occurred before the divergence of Gram-positive and Gram-negative bacteria.

Multiple alignment by using the MACAW program\(^1\) revealed that three motifs are significantly conserved \((p < 10^{-18})\) in all these proteins: (1) a three-residue motif near the N-terminus (PGP in the RhtB subfamily, and PXGP in the LysE subfamily); (2) an aromatic motif that lies ~60 residues from the N-terminus; and (3) an FX\(_7\)LXNPK/HX\(_2\)LX\(_8\)F motif that lies 16–58 residues C-terminal to the second motif (Fig. 1). A highly conserved glycine residue lies 16 residues N-terminal to the second motif, on the edge of the predicted transmembrane segment, and might be part of a three-dimensional flexible hinge that gives mobility to the aromatic residues.

In addition to the three conserved motifs, the RhtB proteins show additional similarity: all are hydrophobic, and their transmembrane segments (predicted by the PHDhtm program\(^2\)) exhibit similar patterns. We propose that they belong to a new, widespread class of functionally important transporters that allow excretion of metabolites from different prokaryotes and archaea.

Acknowledgements

We thank T. P. Turova and R. L. Tatusov for the SEQEDIT manual alignment software, and D. G. Naumoff and M. S. Gelfand for reading the manuscript and for discussions.

References

4 Koonar, R. (1994) EMBO J. 13, 75-83

Vladimir V. Aleshin
State Research Institute of Genetics and Selection of Industrial Microorganisms, Moscow 113565, Russia and Laboratory of Microorganism Biotechnology, Research Institute of Physiolog, Biochemistry and Nutrition of Livestock, Borovsk, Kaluga region, 249010, Russia. Email: petr@bioevol. genebee. msu.su

Natalia P. Zakatueva and Vitaliy A. Livshits
State Research Institute of Genetics and Selection of Industrial Microorganisms, Moscow 113565, Russia.

Pete Jeffs is a freelancer working in Paris, France.