J. Mol. Microbiol. Biotechnol. (2001) 3(1): 113-122.

JMMB Research Article

Organisation and Evolution of the *tol-pal* Gene Cluster

James N. Sturgis

Laboratoire d'Ingénierie des Systèmes Macromoléculaires, UPR 9027, Institut de Biologie Structurale et Microbiologie, 31 Chemin Joseph Aiguier, 13402 Marseille cedex 20, France

Abstract

The genomic context and phylogenetic distribution of the *tol-pal* gene cluster and homologues to its various components have been investigated. The structure of this operon is well conserved across the Gram negative bacteria, and the machine encoded by these genes probably evolved with the appearance of Gram negative bacteria. Since the evolutionary appearance of the operon some species appear to have lost the genes. These bacteria seem to fall into two classes, namely obligate intracellular parasites and bacteria that produce large numbers of outer membrane vesicles. The evolution of the αβ and γ proteobacteria was accompanied by the association of an additional gene (*ybgC*) with the operon. Several coincidences of genomic context argue for an important role of the *tol-pal* operon in cell envelope maintenance. Genes homologous to *tolQ* and *tolR* proved to be very widespread being found throughout the eubacteria, and one example in the archea, this distribution argues for an ancient origin of these genes. The genomic context of these genes often suggests a role in micro-nutrient uptake. Interestingly in all the cases examined the *tolQ* and *tolR* genes or their homologues appear to be present as a pair, with a potential for a tight translational regulation.

Introduction

The *tol-pal* gene cluster coding for seven genes is located at 17 minutes on the +strand of the *E.coli* chromosome (Webster 1991; Vianney *et al.,* 1996). Currently the precise function of the proteins encoded by the gene cluster is unclear, however they appear to be involved in maintaining the integrity of the cell envelope since mutations in many of the genes result in phenotypes that are hypersensitive to certain drugs and to detergents, release periplasmic proteins (Lazdunski *et al.,* 1998) and produce small outer-membrane vesicles (Bernadac *et al.,* 1998). The genes are believed to be organised into an operon containing an internal promotor, the first transcript derived from the genes *ygbC*, *tolQRAB, pal* and *ybgF*; and the second transcript covering the genes *tolB*, *pal* and *ybgF* (Vianney *et al.,* 1996; Müller and Webster 1997).

Received May 22, 2000; revised July 6, 2000; accepted July 10, 2000.
*For correspondence. Email sturgis@ibsm.cnrs-mrs.fr;
Tel. 04 91 16 44 85; Fax. 04 91 71 21 24.

The various proteins produced by the gene cluster are associated with different cellular compartments. The YgbC protein is thought to be a soluble cytoplasmic protein, and a sequence homology has previously been remarked between the soluble 4-hydroxybenzoyl-CoA thioesterase of a soil dewelling *Pseudomonas* species and the *E.coli* YgbC protein (Benning *et al.,* 1998). The proteins TolQ, TolR and TolA are all associated with the cytoplasmic membrane (Derouiche *et al.,* 1995; Lazzaroni *et al.,* 1995). TolQ is an integal membrane protein containing three transmembrane helices and a large cytoplasmic domain between helices 1 and 2. TolR and TolA both contain a single transmembrane helix near the N-terminus and large periplasmic domains. In contrast the proteins coded for by the second operon are all exported. TolB and YbgF are periplasmic proteins (Isnard *et al.,* 1994; Vianney *et al.* 1996) and Pal is a lipoprotein that is associated with the peptidoglycan cell wall (Lazzaroni and Portalier 1992).

A wide variety of interactions between the components of this system have been demonstrated by crosslinking *in vivo* and *in vitro* (Bouveret *et al.,* 1995; 1999; Derouiche *et al.,* 1995; Journet *et al.,* 1999; Cascales *et al.,* 2000), supressor mutation analysis (Clavel *et al.,* 1998; Germon *et al.,* 1998; Ray *et al.,* 2000), and using the yeast double-hybrid system (Walburger and Corda, personal communication). Interactions within the cytoplasmic membrane have been found between TolQ and the other two integral membrane proteins TolA and TolR. Amongst the periplasmic components and domains a large number of possible interactions have been evoked. The homo-dimerisation of both TolB and the periplasmic domain of TolR have been observed, as have the interactions of TolA with TolB, Pal and YbgF, and an interaction between TolB and Pal. A number of interactions have also been demonstrated between the components of the *tol-pal* operon and other cellular components. Both the TolA and TolB proteins interact with porins (Derouiche *et al.,* 1996; Rigal *et al.,* 1997; Clavel *et al.,* 1998), and the TolB protein has also been shown to interact with the major lipoprotein, Lpp (Clavel *et al.,* 1998). Interactions of Pal with the murein sacculus have been demonstrated (Lazzaroni and Portalier 1992), and are associated with a particular sequence motif (Koebnik 1995).

Fractionation and quantification experiments have attempted to localise the components of this system and investigate the stochiomety. Cell fractionation experiments suggest that the products of this gene-cluster are preferentially associated with contact regions between the inner and outer membrane (Guihard *et al.,* 1994). However no clear consensus has been achieved for the stochiometry of the different components, while Pal is present at about 2-3 x 10^3 copies per μm^2 of cell surface area the number of copies per cell varies widely (8000-40000) depending on cell morphology (Cascales and Lloubès, personal communication), the abundance of TolA and TolR have been measured as about 600 (Levengood *et al.,* 1991) and 2500 (Müller *et al.,* 1993) copies per cell respectively,

but without reference to the cell surface area or morphology. These results would nevertheless appear to indicate an excess of the outer-membrane associated Pal over the inner-membrane associated components TolA and TolR. This stochiometry is entirely consistent with the data on the additional transcription of the second part of the operon from an internal promoter (Müller and Webster 1997).

Recently crystallographic structures of two different parts of the system have been published, the C-terminal domain of TolA (TolA3) in interaction with g3p (Lubkowski *et al.,* 1999), and the TolB protein (Abergel *et al.,* 1999; Carr *et al.,* 2000). The C-terminal domain of TolA is composed of a globular α/β domain containing 3 α helices which cover a 3 stranded β sheet, with extensions into a bundle containing 2 helices and an aperiodic strand. The structure of TolA3 creates a depression in which is bound the g3p domain of the fusion protein used for the structure determination. The structural relationship between the g3p associated form and the native form (or forms) of the TolA3 domain is as yet unclear. The TolB protein contains two distinct domains, a C-terminal 6 bladed β propeller domain and an N-terminal α/β domain. The α/β domain contains 2 α helices which cover a 5 strand β sheet. The two structural domains associate via a relatively restricted and possibly mobile contact area. Analysis of the sequence of TolB and homologous proteins identified in other bacterial genomes gave rise to the suggestion of a role for this protein in peptidoglycan metabolism (Abergel *et al.,* 1999).

The integral membrane proteins of the *tol-pal* system (TolQ, TolR and TolA) are related to another group of integral membrane proteins ExbB, ExbD and TonB (Eick-Helmerich and Braun 1989; Kampfenkel and Braun 1992; 1993; Müller *et al.,* 1993; Vianney *et al.,* 1994). These latter proteins are involved in siderophore and cobalamin import across the outer membrane. Indeed the similarities between the ExbBD and TolQR proteins are sufficient to allow some interchange of components and cross talk between the two systems, thus the effects of ExbB or ExbD

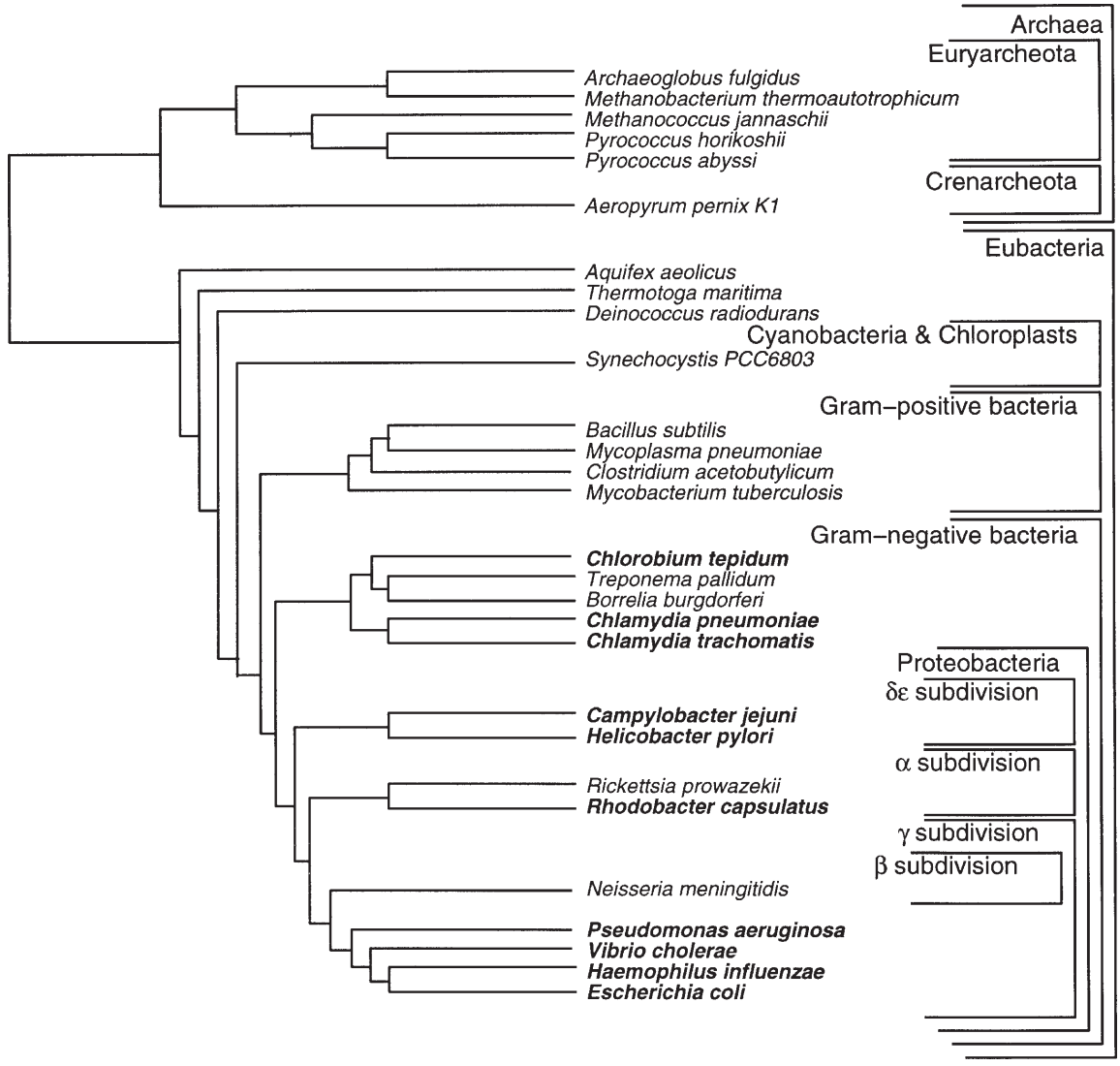

Figure 1. Phylogenetic tree based on rRNA sequences showing the relationship between the various organisms that have (in bold) or do not have a *tol-pal* gene cluster orthologous to that of *E. coli.* The tree is designed to show the branching pattern, but the branch lengths have no significance.

loss of function mutations are much more pronounced in a TolQ or TolR deficient strain (Braun and Hermann 1993; Larsen *et al.,* 1999). The TonB protein appears to be analogous to the TolA protein mediating an interaction between the inner-membrane components (ExbB and ExbD) and the outer-membrane components, which in the case of TonB are outer-membrane siderophore and vitamin B12 receptors.

In view of the large number of interactions between the different proteins of the *tol-pal* system it seems likely that the various proteins produced by this gene cluster form part of a multi-protein complex, and further that the presence of the gene cluster in the *E.coli* genome reflects a functional requirement for the co-ordinated regulation of expression and the control of component stochiometry. We have therefor investigated, and describe here, the phylogenetic distribution of orthologues to this gene cluster, and its various component genes in the hope of better understanding its biological role.

In previous work some individual components of the *tol-pal* gene cluster have been identified in various bacteria (Abergel *et al.,* 1999), or groups of genes for example in *Pseudomonas aeruginosa* (Dennis *et al.,* 1996; Duan *et al.,* 2000) or *Vibrio cholerae* (Heilpern and Waldor 2000). However, previous authors have investigated neither the context of these genes nor their phylogenetic distribution, investigations which are now becoming possible with the explosion of microbial genomic sequence information.

Results and Discussion

Distribution of the *tol-pal* Gene Cluster

In order to investigate the distribution of the *tol-pal* gene cluster we searched all complete and a variety of unfinished genomes for gene clusters containing open reading frames coding for proteins homologous to those coded for by the *E.coli tol-pal* cluster. The sequences of the cluster which proved the most useful in this search were those of TolQ and Pal, which were easy to identify in many databases and had well conserved regions, while other sequences proved virtually useless, such as TolA, which proved very poorly conserved and almost impossible to find, except with reference to its context.

We thus searched for clusters of predicted open reading frames containing several genes of the *tol-pal* operon on the same strand. We were able to find such gene clusters in genomes for 10 different species spanning the Gram-negative bacteria, the phylogenetic distribution is illustrated in Figure 1. It should be noted that the relatively low number of clusters found is due to the fact that the size of the *tol-pal* gene cluster (6 kbases in *E.coli*) is much larger than the contigs available for many incomplete genomes. Thus, the absence in many sequenced but unassembled genomes is simply due to uncertainties about the organisation on the kbase scale. However, the gene cluster was not found in any archaea (Bult *et al.,* 1996; Smith *et al.,* 1997; Klenk *et al.,* 1997; Kawarabayasi *et al.,* 1998; 1999), in any Gram positive eubacteria (Fraser *et al.,* 1995; Himmelreich *et al.,* 1996; Kunst *et al.,* 1997; Cole *et al.,* 1998), or in *Aquifex aeolicus* (Deckert *et al.,* 1998), *Thermotoga maritima* (Nelson *et al.,* 1999), *Deinococcus radiodurans* (White *et al.,* 1999) or *Synechocystis* PCC6803 (Kaneko *et al.,* 1996) genomes. Within the completely sequenced Gram negative eubacterial genomes, an equivalent *tol-pal* gene cluster was not found in either of the spirochete genomes (*Treponema pallidum* (Fraser *et al.,* 1998) and *Borellia burgdorferi* (Fraser *et al.,* 1997; Casjens *et al.,* 2000)) nor was it found in the genomes of *Rickettsia prowazeki* (Andersson *et al.,* 1998), or *Neisseria meningitidis* (Tettelin *et al.,* 2000; Parkhill *et al.,* 2000b). However the cluster was found across the gram negative bacteria, outside the proteobacteria (*Chlorobium tepidum* (TIGR), *Chlamydia pneumoniae* (Kalman *et al.,* 1999; Read *et al.,* 2000) and *Chlamydia trachomatis* (Stephens *et al.,* 1998; Read *et al.,* 2000)) and within the proteobacteria in the α (*Rhodobacter capsulatus* (University of Chicago)), γ (*E. coli* (Blattner *et al.,* 1997), *Haemophillus influenzae* (Fleischmann *et al.,* 1995), *Vibrio cholerae* (TIGR)) and $\delta\varepsilon$ (*Helicobacter pylori* (Tomb *et al.,* 1997; Alm *et al.,* 1999), *Campylobacter jejuni* (Parkhill *et al.,* 2000a)) subdivisions.

The absence of the cluster in Gram positive bacterial genomes is easy to rationalise based on its implication in maintenance of the outer-membrane integrity. The absence in some Gram-negative genomes shows that the cluster is not universal though widespread, it is perhaps interesting that the various gram negative bacteria where it was not found are intracellular parasites.

Structure of the *tol-pal* Operon

The structure of the different clusters identified was remarkably constant, and is illustrated in Figure 2. The various genes and open reading frames in the clusters are labelled according to the annotations found in the sequence databases and the most homologous proteins identified - where this gave a clear indication of function, except that those orthologues to *tolQ* and *tolR* which were annotated as, or most homologous to, *exbB* or *exbD* have been renamed appropriately. The structure of the operon is remarkably stable across a wide phylogenetic space. The *tolQ*, *tolR*, *tolB* and *pal* genes had in general been annotated and are easy to recognise. In most cases an unannotated gene occupies the space between *tolR* and *tolB*. In *H. pylori* the situation is a little confused, since in the 26695 genome there are two open reading frames annotated in this region, while in the case of the J99 strain there is only one open reading frame, which shows some slight homology to the *tolA* gene. It is possible that the frame shift responsible for the difference in the sequences is erroneous and that the 26695 genome also contains a single *tolA* gene between the genes coding for TolR and TolB. In the gene clusters collected from the α and γ proteobacteria, there was an equivalent of the *ybgC* gene upstream of the *tolQ* gene. Furthermore, in each cluster, with the exception of *H. influenzae*, there was a gene after *pal* that appeared to belong to the *tol-pal* gene cluster, equivalent to *ybgF* in *E. coli*. In the genomes of the *Chlamydia* species this gene has been identified as a homologue to *amiA* of *E. coli*, a probable N-acetyl-muramoyl-L-alanine amidase.

In the 10 different gene clusters examined the predicted inter-cistronic gaps are often very short, with only one third of gaps larger than the 35 base ribosomal footprint, the predicted inter-cistronic gap sizes are tabulated in Table 1. This compactness is particularly

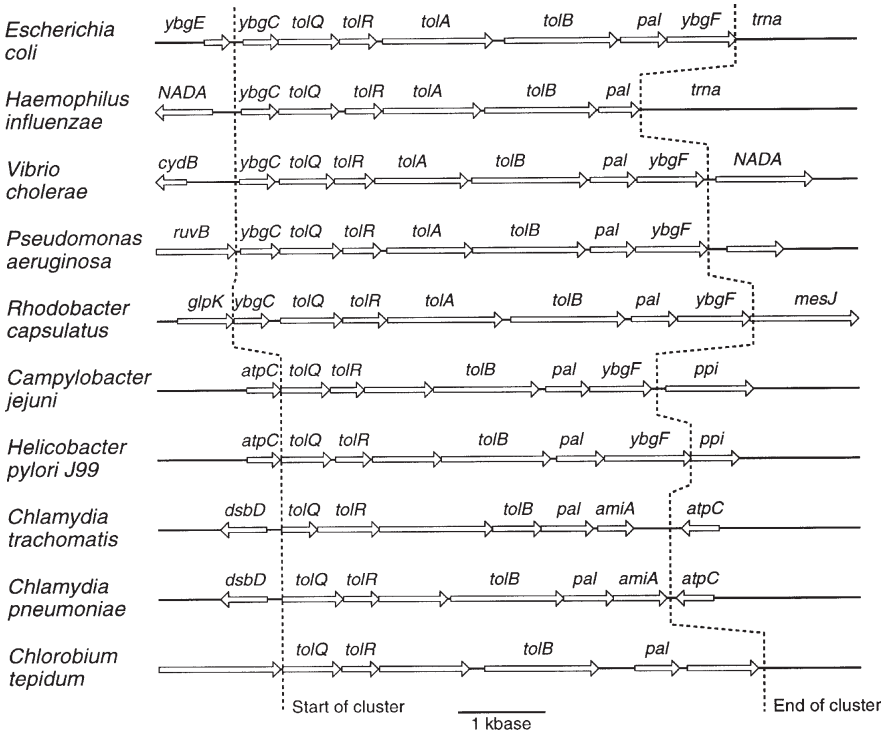

Figure 2. Organisation of the *tol-pal* gene clusters identified. The size, order and context of the genes are indicated for the 10 identified clusters. Genes shown as arrows are labelled according to their known functions or homologies, unlabelled genes do not show particular homology to genes of known or inferred function.

noticeable in *C. jejuni* (average gap 19.6 bases) and the *Chlamydia* species (average gaps 6.4 and 3 bases). Viewing the table in the orthogonal direction the gaps *tolQ-tolR*, and *pal-ybgF* appear to be consistently small across the different genomes (average gaps 17.8 and 19.2 bases respectively). This compactness would argue for a tight and concerted regulation, at least in certain species, of the translation of the different genes from the messenger RNA, and thus provides evidence for the biological importance of a co-ordinated regulation of the genes in the *tol-pal* operon.

Annotation of the *tol-pal* Cluster

The annotation of the *tolA* gene without reference to its context proved very difficult. Indeed the majority of *tolA* genes found in sequenced genomes were not annotated. This difficulty was due to very poor conservation between even closely related species. Thus the *tolA* gene represents a highly divergent gene with a very strongly species-dependant sequence. For example comparing the various gene products between *E.coli* and *H.influenzae* while the level of sequence identity for members of the cluster is 55 ± 6% the level of sequence identity for the TolA proteins is significantly lower, only 32%. This low level of conservation, relative to that of the other gene products, appears to be a feature of the TolA protein and is independent the pair of sequences used for the comparison. Nevertheless, this sequence divergence is associated with a strongly conserved context, and an apparently necessary function in the context of the *tol-pal* cluster.

Table 1. Structure of the tol-pal Operon: Predicted Inter-Cistronic Gap Sizes in Nucleotides

Species	*YbgC tolQ*	*tolQ tolR*	*tolR tolA*	*tolA tolB*	*tolB pal*	*pal ybgF*
Cb. tepidum	-	14	15	172	412	84
Ch. pneumoniae	-	-4	3	31	-4	-11
Ch. trachomatis	-	2	-1	4	-11	46
Ca. jejuni	-	8	2	9	76	3
H. pylori J99	-	51	18	-4	66	7
Rb. casulatus	127	6	11	93	69	8
P. aeruginosa	4	25	65	-1	55	12
V. cholerae	46	2	17	41	31	15
H. influenzae	16	71	15	41	23	-
E. coli	-4	3	64	132	34	9

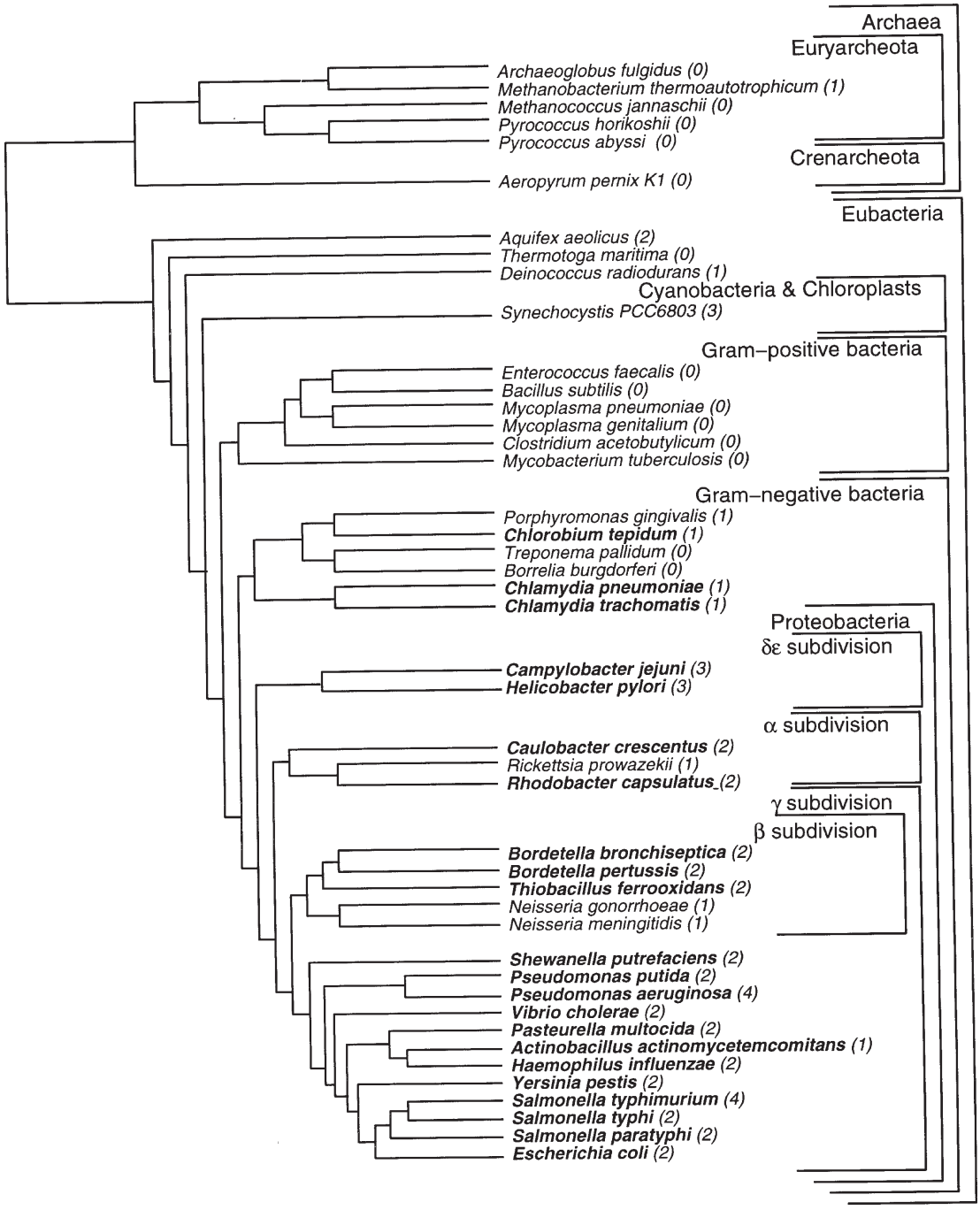

Figure 3. Phylogenetic tree based on rRNA sequences showing the relationship between the various organisms and groups mentioned in the text. Those species marked in bold typescript have (or almost certainly have) a *tol-pal* gene cluster orthologous to that of *E. coli*. The numbers in parenthesis after each species indicates the number of distinct gene pairs homologous to *tolQ* and *tolR* found in the available sequences.

The pair of genes *tolQ* and *tolR* were often annotated in the databases as *exbB* or *exbD* homologues, and indeed in a number of cases appeared more similar to these than to *tolQ* and *tolR*. However, in view of their context within a *tol-pal* gene cluster, it is clear that they are orthologues to *tolQ* and *tolR*. This difficulty of annotation was investigated further to see if some internal sequence signature could differentiate *tolQR* gene pairs in the context of a *tol-pal* cluster from other gene pairs. Unfortunately no such signature was found.

In the 10 *tol-pal* operons identified 9 contained a gene after *pal* that apparently belongs to the cluster. This gene, *ybgF* in the case of *E. coli*, is of unknown function, however the gene occupying the equivalent position in the operons from *C. trachomatis* and *C. pneumoniae* is probably an N-acetyl-muramoyl-L-alanine amidase. Unfortunately it has

proved impossible to find a convincing alignment of the 9 sequences to reinforce the idea of a common evolutionary origin, rather it appears that all the proteins except those derived from the *Chlamidia* genomes are related but different from the *Chlamydia* proteins which are probably N-acetyl-muramoyl-L-alanine amidases. It is of course pertinent that once again a relationship is found between genes of the *tol-pal* operon and cell wall metabolism.

Context of the *tol-pal* Cluster

Despite the relative constancy found for the arrangement within the gene cluster, for example no changes in the gene order, the genomic context varied considerably. Thus while in *E. coli* and *H. influenzae* the genes are followed by genes for tRNA's, this is not the case in other genomes where the operon is found between predicted open-reading frames. In general there seems to be little consistency of function for the genes surrounding the operon, they are not for example all involved in cell wall metabolism. A striking similarity of context is however evident in comparing the *Campylobacter* and *Helicobacter* operons with those from the *Chlamydia* species where in each case the cluster is located between the genes for the ATPase ε subunit *atpC* and the gene for a protein involved in the catalysis of protein folding prolyl-peptidyl isomerase (*ppi*) or disulphide bond isomerisation (*dsbD*), respectively.

Probable Distribution of the *tol-pal* Operon in Other Bacteria

A closer examination of the various genes localised in sequences derived from other proteobacteria indicated that at least the first six genes (*ybgC* to *pal*) are probably present in an orthologous cluster in the genomes a wide variety bacteria from the β and γ subdivisions (see Figure 3). In these cases homologous genes and pairs were found in the currently available sequences that were entirely consistent with the presence of an orthologous cluster, i.e. if a reading frame adjacent to a homologous gene could be identified it coded a protein homologous to the adjacent gene in the *tol-pal* operon. Unfortunately the last gene of the cluster (*ybgF*) and *tolA* proved too variable to allow reliable and consistent identification. On the basis of this search the *tol-pal* gene cluster is postulated to exist in the genomes of: *Caulobacter crescentus*, *Bordetella bronchiseptica*, *Bordetella petussis*, *Thiobacillus ferrooxidans*, *Shewanella putrefaciens*, *Pseudomonas putida*, *Pasturella multocida*, *Actinobacillus actinomycetocomitans*, *Yersinia pestis*, and three *Salmonella* species, *typhi*, *typhimurium* and *paratyphi*. However, it seems unlikely, on the basis of the current sequence information, that the cluster will be found in *Porphymonas gingivalis*, or *Neisseria gonorrhoeae*.

Thus a gene cluster orthologous to the *tol-pal* cluster of *E. coli* has either been found, or seems very likely to exist, in almost all Gram negative bacteria, with a number of notable exceptions. In contrast no evidence for an orthologous cluster has been found outside this group. Those species in which a *tol-pal* gene cluster would be expected, based on their phylogenetic position as gram negative bacteria, but appear not to contain such a cluster are *P. gingivalis*, *T. pallidum*, *B. burgdorferi*, *R. prowazeki*, *N. gonorrhoeae* and *N. meningitidis*.

This extended search also clarified the structure of the operon. Within the α, β and γ subdivisions of the proteobacteria the *tol-pal* operon appears to contain an orthologue of the *ybgC* gene, while outside this grouping the cluster seems to begin with the *tolQ* gene.

Distribution of Homologous Genes Outside a *tol-pal* Operon

Since our searches had turned up a number of homologous genes that appeared in various contexts, the phylogenetic distribution and genomic context of these homologous genes outside the context of the *tol-pal* gene cluster was also examined.

TolQ/R Homologues

The most widely distributed were *tolQ* and *tolR* homologues, such as the *exbB* and *exbD* genes of *E.coli*. Homologues of the *tolQ* and *tolR* genes were much more widely distributed than the entire *tol-pal* operon, even amongst completely sequenced genomes. Such a pair of homologous genes were found in species that contained the *tol-pal* gene cluster, for example the *exbB* and *exbD* gene pair in *E. coli*, and in species in which the *tol-pal* gene cluster appears to be absent, for example *R. prowaseki* or *Neiserria meningditidis*. In many genomes multiple copies of this pair of genes were found outside the context of a *tol-pal* gene cluster, thus for example in *P. aeruginosa* three pairs of genes were found, for the various species searched this information is noted in Figure 3. Phylogenetically the gene pair was found throughout the eubacteria, with the exception of two recognised groups the gram positive bacteria and the spirochetes. In all cases the genes were present as a pair, in the same order *tolQ* followed by *tolR*. Interestingly an example was also found in an archae genome, that of *Methanococcus thermoautotrophicum*, though the homology of the putative TolR equivalent was very weak. This pair of genes was only distantly related to the majority of eubacterial genes. It is difficult to assess unequivocally whether this gene pair is a vestige of ancient origin or the result of horizontal gene transfer between archea and eubacteria. However two lines of evidence would argue in favour of a horizontal transfer: first the absence of the gene pair in other members of the archea, and second the similarity between the *Methanococcus* TolQ like sequence and that of some eubacterial TolQ homologues notably that of *Porphymonas gingivalis*.

The wide distribution of this gene pair suggests both an ancient origin, probably within the eubacteria, and an important function, or functions. However our current level of understanding does not permit the identification of a unique and general physiological role for this group of proteins.

The context of homologues to TolQ and TolR outside the tol-pal gene cluster was variable. In many cases (about 50%) the genes were found associated with homologues to *tonB*, or with other genes involved in micro-nutrient uptake, such as a transporter homologous to *corA* (a gene involved in magnesium transport) in *H. pylori*. Indeed in *C. jejuni* multiple copies were found each associated with *tonB* homologues. In most of the remaining cases the genes were isolated, either by regions of non-coding DNA or by

changes in the coding strand, as for example in *R. prowazekii* or *E. coli*, making any functional inferences very tentative.

The preponderance of associations with homologues to TonB suggests in many cases the TolQ/R proteins will be associated with a transperiplasmic protein like TonB (or TolA). This association appears to be able to function as an energy transmission mechanism deriving energy from the cytoplasmic membrane proton-motif force (Larsen *et al.,* 1999; Cascales *et al.,* 2000). Furthermore the associations with other proteins involved in micro-nutrient capture suggest a frequent involvement in such processes, as is known for the ExbB/D proteins of *E. coli*.

It is noteworthy that in none of the cases where TolQ and TolR homologues were found outside the context of a *tol-pal* cluster was a gene showing homology to *ybgC* found. This point strengthens the idea that this gene is an integral part of the *tol-pal* operon and is associated with the function of these genes.

Other Proteins

Outside the example of *tolQ* and *tolR* two different cases were examined, an equivalent to *ybgC* was found in *H. pylori* and *C. jejuni*, equivalents of *tolB* and *pal* were found in *R. prowazekii.* The *ybgC* homologues were associated in the case of the *H. pylori* genomes with other genes known to be involved in cell envelope stability, notably *murD* and *mraY*, however in *C. jejuni* the gene is associated with a small putative coding region of 32 residues. The association in *H. pylori* of a gene from the *tol-pal* gene cluster with genes involved in cell wall synthesis would appear to reinforce the idea that this gene forms part of the *tol-pal* cluster and that the *tol-pal* system is necessary for cell envelope maturation. However the possible role of a putative acyl-thioesterase or acyl-transferase in this process remains a subject of investigation.

In the case of *R. prowazekii* genes similar to *tolB* and *pal* were found. In each case the gene is isolated from other genes by long non-coding regions and changes in coding strand. This renders the situation of the Tol-Pal complex a little ambiguous in this bacteria, since four of the five core genes (*tolQRB* and *pal*) of the cluster can be identified but unlike other proteobacteria in which they are conserved in a cluster the genes are distributed across the genome. Futhermore the lack of an identified TolA homologue is probably more related to our lack of ability in defining a characteristic signature of this protein than anything else. Thus it seems probable that the genes of the *tol-pal* cluster are present but distributed across the genome. It is of course at this stage impossible to say whether this lack of coherence is associated with a loss of function or not.

In the cases of the bacteria that would be expected to contain a *tol-pal* cluster based on their phylogenetic location, as Gram negative bacteria, but appear not to, three classes are observed. In the spirochetes, *Treponema palidum* and *Borellia burgdorferii*, no homologues of members of the *tol-pal* gene cluster were found. In *Porphymonas gingivalis*, *Neisseria gonorrhoeae* and *Neisseria meningitidis*, homologues to the *tolQ* and *tolR* genes appear to be present, but no homologues to other members of the gene cluster were found. Finally in *Rickettsia prowasekii* the genes appear to be present but distributed across the genome. Apart from these exceptions the complete and partially complete genomes examined indicate the presence of a conserved *tol-pal* gene cluster.

Conclusions

Evolutionary Origin of the *tol-pal* Gene Cluster

The distribution of the *tol-pal* gene cluster provides information that is of interest in considering the evolution of this membrane-associated protein complex. The original starting point for the evolution of this system would appear to be a *tolQ/R* gene pair, present in the ancestral eubacteria, or possibly in view of the detection of a similar gene-pair in *M. thermautotrophicum* even before the emergence of the eubacteria. This original gene pair was subsequently embellished during the development of Gram negative bacteria, but after the separation of the cyanobacteria, and the unusual eubacteria such as *A. aeolicus*, *D. radiodurans* or *T. maritima*, the sequenced genomes of these organisms appear to contain *tolQ* and *tolR* homologues but lack associations with homologues of any of the other members of the *tol-pal* gene cluster.

It is difficult to place the evolution of the *tol-pal* gene cluster relative to the evolution of the Gram positive bacteria since this group of bacteria lack both homologues to the *tol-pal* gene cluster and even homologues to the *tolQ/tolR* gene pair. However by the time the first Gram negative bacteria had evolved it seems probable that a functional *tol-pal* gene cluster had been established in the genome, this development prior to the evolution of the proteobacteria is indicated by the detection of intact *tol-pal* gene clusters in two *Chlamydia* species genomes and the genome of *Chlorobium tepidum*. Since the first appearance of the tol-pal system the gene cluster has been lost in a number of groups, notably the *Spirochaetes*, *Porphymonas gingivalis*, and the two species of *Neisseria*. In the case of *R. prowazekii* the genes appear to be on the road to loss as the gene cluster has lost its coherence and we would predict the *tol-pal* system is either no longer, or only very rarely, functional in this bacterium in view of the loss of stochiometry control that would be expected to accompany the loss of operon coherence.

Since the appearance of the proteobacteria one major change in the *tol-pal* gene cluster has occurred. Some time after the separation of the $\delta\varepsilon$ proteobacteria from the $\alpha\beta\gamma$ proteobacteria this later group added the *ybgC* homologue to the gene cluster, as the first gene. The YbgC protein appears to be an acyl transferase, based on the strong homology to such proteins and the conservation of the active site residues (Benning *et al.,* 1998). However the importance of an acyl transferase within a Tol-Pal complex or the tol-pal gene cluster, especially as the first gene of a tightly coupled series in a polycistronic message remains a mystery. Is the protein required for the acylation of another component of the system, as for example in the production of haemolysin (Stanley *et al.,* 1998) or is a trans-acylase activity required as an initial activity of the *tol-pal* molecular machine? Preliminary experiments have not yet provided a clue as to the role of this protein (Sun and Webster 1987), and despite the large number of interactions catalogued between the products of the operon none involve YbgC.

Functional Role of the *tol-pal* Gene Cluster

The wide distribution of the *tol-pal* gene cluster, the conservation of the gene organisation, and the indications of tight translational regulation all suggest an important physiological role for the products of these genes in Gram negative bacteria. The tight regulation of the different components of this system is perhaps a result of the large number of interactions that have been observed between the different components. Though the precise role of these genes is unknown the distribution of the gene cluster, and the physiological peculiarities of these Gram negative bacteria that lack the genes all point to a role in the maintenance of outer membrane integrity. It is noteworthy that those organisms lacking, or postulated to lack, a functional Tol-Pal system fall into two groups. Bacteria that appear to be obligate intracellular parasites (*B.burgdorferi*, *T.pallidum* and *R. prowazekii*) on the one hand, and bacteria known for an outer membrane instability that leads to the production of large numbers of outer membrane vesicles (*P. gingivalis*, *N. meningditidis* and *N. gonorrhoae*) (Zhou *et al.,* 1998; Dorward *et al.,* 1989), a characteristic also found in *E. coli* mutants lacking a functional Tol-Pal system (Bernadac *et al.,* 1998).

The pair of genes homologous to *tolQ* and *tolR* are more widespread than the entire *tol-pal* cluster. In certain cases it is probable that this pair of genes are involved in the uptake of iron and/or cobalamine, analogous to the *tonB* system of *E. coli*. However in those species possessing many analogous pairs it remains far from clear what other roles (if any) this pair of genes can perform, do these bacteria contain redundant pairs, or are the extra pairs of genes involved in some other process(es)?

The Evolution of TolA

The lack of conservation of the TolA sequence is intriguing. Particularly in view of the large number of interactions of this protein, both within the tol-pal complex with TolQ, TolB, Pal, YbgF (Derouiche *et al.,* 1995; Germon *et al.,* 1998; Cascales *et al.,* 2000; Walburger and Corda, personnel comunication) and outside the context of the Tol-Pal complex with porins (Derouich *et al.,* 1996; Rigal *et al.,* 1997). All the potential partners appear much more strongly conserved than the TolA protein. Thus paradoxically the least conserved protein appears to be involved in the most interactions, interacting with both the cytoplasmic membrane components and the periplasmic components. Furthermore the regions and sequences involved in these interactions are no more conserved than the rest of the protein, while the protein is one of the most important for outer membrane stability.

This paradoxical situation can perhaps be understood in view of the wide parasitism of TolA. This protein is used by the colicins (Lazdunski *et al.,* 1995) many viruses (Reichmann and Holliger 1997) as a co-receptor to gain entry into *E. coli*. It is also probably used similarly by other bacteriophages to gain entry into other species, for example the entry of the CTX phage into *V. cholerae* (Heilpern and Waldor. 2000). It thus seems possible that there is a selective advantage to mutations in the *tolA* gene - as this will provide immunity to both colicins and filamentous phages - and further that these mutations might aid speciation by preventing gene transfer by these phages.

The TolA protein is thus a strongly species specific protein, and perhaps even a strain specific protein, (there are significant strain specific differences in those cases where multiple sequences are known - *Chlamydia pneumoniae*, *Chlamydia trachomatis* and possibly *Helicobacter pylori*) that exists in a context conserved throughout the gram negative bacteria. Clearly the variability of the *tolA* gene deserves further investigation in the hope of better understanding the selective pressures on this protein.

Experimental Procedures

Computational Methods

Sequences were retrived from blast searches at the NCBI (http://www.ncbi.nlm.nih.gov/BLAST/) WIT2 at the university of Oklahoma (http://dna1.chem.ou.edu:8080/WIT2/), TIGR (http://www.tigr.org) and Pedant (http://pedant.mips.biochem.mpg.de) sites, or at websites accessed via these, against the sequences, or portions of the sequences, of proteins coded by the *E. coli tol pal* cluster or their orthologues. Candidate homologous genes with their contexts were retrieved and examined, in the case of unannotated sequences potential open reading frames were identified using the NCBI ORFfinder programme. Multiple sequence alignments used to identify characteristic signatures of the different proteins were performed with the clustal W programme (Thompson *et al.,* 1994). The phylogenetic trees displayed in Figures 1 and 3 are derived from the rRNA sequence based trees available through the ribosomal database project at Michigan state university (http://www.cme.msu.edu/RDP/).

Acknowledgements

This work would not have been possible without the assistance and encouragement of Roland Lloubès and Danielle Cavard, who have undertaken - through many discussions and answers to my sometimes naive questions - to educate me in matters of bacteriology and the wonders of the Tol-Pal system. I would also like to thank Chantal Abergel for many useful discussions and Claude Lazdunski for introducing me to this subject.

Some preliminary sequence data were obtained from the Institute for Genomic Research website at http://www.tigr.org. The sequencing of *Chlorobium tepidum* and *Vibrio cholerae* by TIGR was made possible by funding from the US-DoE and NIAID repectively

References

Abergel C., Bouveret E., Claverie J., Brown K., Rigal A., Lazdunski C. and Bénédetti H. 1999. Structure of the *E. coli* TolB protein determined by MAD methods at 1.95 Å resolution. Structure. 7: 1291-1300.

Alm RA, Ling LS, Moir DT, King BL, Brown ED, Doig PC, Smith DR, Noonan B, Guild BC, deJonge BL, Carmel G, Tummino PJ, Caruso A, Uria-Nickelsen M, Mills DM, Ives C, Gibson R, Merberg D, Mills SD, Jiang Q, Taylor DE, Vovis GF, Trust TJ. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori.* Nature. 397: 176-180.

Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski RM, Naslund AK, Eriksson AS, Winkler HH, Kurland CG. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature. 396: 133-140.

Benning M.M., Wesenberg G., Lui R., Taylor K.L., Dunaway-Mariano D. and Holden H.M. 1998. The three dimensional structure of 4-hydroxybenzoyl-CoA thioesterase from *Pseudomonas* sp. strain CBS-3 J. Biol. Chem. 273: 33572-33579.

Bernadac A., Gavioli M., Lazzaroni J-C., Raina S. and Lloubès R. 1998. *E. coli tol-pal* mutants form outer membrane vesicles. J. Bacteriol. 180: 4872-4878.

Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. 1997. The complete genome sequence of *Escherichia coli* K-12. Science. 277: 1453-1474.

Bouveret E., Bénédetti H., Rigal A., Loret E. and Lazdunski C. 1999. *In vitro* characterization of the peptidoglycan associated lipoprotein Pal. - peptidoglycan and Pal - TolB interactions. J. Bacteriol. 181: 6306-6311.

Bouveret E., Derouiche R., Rigal A., Lloubès R. Lazdunski C. and Bénédetti H. 1995. Peptidoglycan associated lipoprotein - TolB interaction J. Biol. Chem. 270: 11071-11077.

Braun V. and Herrmann C. 1993. Evolutionary relationship of uptake of

biopolymers in *E. coli*: cross complementation between the TonB-ExbB-ExbD and TolA-TolQ-TolR proteins. Mol. Microbiol. 8: 261-268.

Braun V., Gaisser S., Herrmann C., Kampfenkel K., Killmann H. and Traub I. 1996. Energy coupled transport across the outer membrane of *E. coli*: ExbB binds ExbD and TonB *in vitro*, and leucine 132 in the periplasmic region and aspartate 25 in the transmembrane region are important for ExbD activity. J. Bacteriol. 178: 2836-2745.

Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb JF, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NSM, Venter JC. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science. 273: 1058-1073.

Carr S., Penfold C.N., Bamford V., James R. and Hemmings A.M. 2000. The structure of TolB, an essential component of the tol-dependent translocation system and its protein-protein interaction with the translocation domain of colicin E9. Structure. 8: 57-66.

Cascales, E., Gavioli, M., Sturgis, J.N., and Lloubès, R. 2000. Proton motive force drives the interaction of the inner membrane TolA and the outer membrane Pal proteins in *Escherichia coli*. Mol. Microbiol. 38: In press.

Casjens S, Palmer N, van Vugt R, Huang WM, Stevenson B, Rosa P, Lathigra R, Sutton G, Peterson J, Dodson RJ, Haft D, Hickey E, Gwinn M, White O, Fraser CM. 2000. A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi.* Mol. Microbiol. 35: 490-516.

Clavel T., Germon P. Vianney A. Portallier R. and Lazzaroni J-C. 1998. TolB protein of *E. coli* K-12 interacts with the outer membrane peptidoglycan associated proteins Pal, Lpp and OmpA. Mol. Microbiol. 29: 359-367.

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Barrell BG, *et al.,* 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. 393: 537-544.

Deckert G, Warren PV, Gaasterland T, Young WG, Lenox AL, Graham DE, Overbeek R, Snead MA, Keller M, Aujay M, Huber R, Feldman RA, Short JM, Olsen GJ, Swanson RV. 1998. The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus.* Nature. 392: 353-358.

Dennis J. Lafontaine E. and Sokol P. 1996. Identification and characterization of the *tolQRA* genes of *Pseudomonas aeruginosa*. J. Bacteriol. 178: 7059-7068.

Derouiche R., Bénédetti H., Lazzaroni J-C., Lazdunski C. and Loubès R. 1995. Protein complex within *E. coli* inner membrane J. Biol. Chem. 270: 11078-11084.

Derouiche R., Gavioli M., Bénédetti H., Prilipov A., Lazdunski C. and Lloubès R 1996. Tol A central domain interacts with *Escherichia coli* porins. EMBO J. 15: 6408-6415.

Dorward D.W., Garon C. F. and Judd R. C. 1989. Export and intercellular transfer of DNA via membrane blebs of *Neisseria gonorrhoeae*. J. Bacteriol. 171: 2499-2505.

Duan K, Fafontaine E. R., Majumdar S and Sokol P. A. 2000. Reg A, Iron and growth phase regulate expression of the *Pseudomonas aeruginosa tol-oprL* gene cluster. J. Bacteriol. 182: 2077-2087.

Eick-Helmerich K. and Braun V. 1989. Import of biopolymers into *E. coli*: nucleotide sequences of the *exbB* and *exbD* genes are homologous to those of *tolQ* and *tolR* genes respectively. J. Bacteriol. 171: 5117-5126.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EFKerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, *et al.,* 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science. 269: 496-512.

Fraser CM, Casjens S, Huang WM, Sutton GG, Clayton R, Lathigra R, White O, Ketchum KA, Dodson R, Hickey EK, Gwinn M, Dougherty B, Tomb JF, Fleischmann RD, Richardson D, Peterson J, Kerlavage AR, Quackenbush J, Salzberg S, Hanson M, van Vugt R, Palmer N, Adams MD, Gocayne J, Venter JC, *et al.,* 1997. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi.* Nature. 390: 580-6.

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, *et al.,* 1995. The minimal gene complement of *Mycoplasma genitalium*. Science. 270: 397-403.

Fraser CM, Norris SJ, Weinstock GM, White O, Sutton GG, Dodson R, Gwinn M, Hickey EK, Clayton R, Ketchum KA, Sodergren E, HardhamJM, McLeod MP, Salzberg S, Peterson J, Khalak H, Richardson D, HowellJK, Chidambaram M, Utterback T, McDonald L, Artiach P, Bowman C, Cotton MD, Venter JC, *et al.,* 1998. Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. Science. 281: 375-388.

Germon P., Clavel T., Vianney A., Portallier R. and Lazzaroni J-C. 1998. Mutationlal analyses of the *E. coli* K-12 TolA N-terminal region and characterisation of its TolQ-interacting domain by genetic suppression. J. Bacteriol. 180: 6433-6439.

Guihard G., Boulanger P., Bénédetti H., Lloubès R., Besnard M. and Letellier L 1994. Colicin A and Tol proteins involved in its translocation are preferentially located in the contact sites between the inner and outer membrane of *E. coli* cells. J. Biol. Chem. 269: 5474-5880.

Heilpern A.J. and Waldor M.K. 2000. CTXϕ infection of *Vibrio cholerae* requires the *tolQRA* gene products. J. Bacteriol. 182: 1739-1747.

Higgs P., Meyers P. and Postle K. 1998. Interactions of the TonB dependant energy transduction complex ExbB and ExbD form homo-multimers. J. Bacteriol. 180: 6031-6038.

Himmelreich R, Hilbert H, Plagens H, Pirkl E, Li BC, Herrmann R. 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. Nucl. Acids Res. 24: 4420-4449.

Isnard M., Rigal A., Lazzaroni J-C., Lazdunski C. and Lloubès R 1994. Maturation and localization of the TolB protein required for colicin import. J. Bacteriol. 176: 6392-6396.

Journet L., Rigal A., Lazdunski C. and Bénédetti H. 1999. Role of TolR N-terminal, central and C-terminal domains in its dimerization and interaction with TolA and TolQ. J. Bacteriol. 181: 4476-4484.

Kalman S, Mitchell W, Marathe R, Lammel C, Fan J, Hyman RW, Olinger L, Grimwood J, Davis RW, Stephens RS. 1999. Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. Nat. Genet. 21: 385-389.

Kampfelkel K. and Braun V. 1992. Membrane topology of the *E. coli* ExbD protein. J. Bacteriol. 174: 5485-5487.

Kampfelkel K. and Braun V. 1993. Membrane topology of the *E. coli* ExbB protein. J. Biol. Chem. 268: 6050-6057.

Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirosawa M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M, Tabata S. 1996. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res. 3: 109-136.

Kawarabayasi Y, Hino Y, Horikawa H, Yamazaki S, Haikawa Y, Jin-no K, Takahashi M, Sekine M, Baba S, Ankai A, Kosugi H, Hosoyama A, Fukui S, Nagai Y, Nishijima K, Nakazawa H, Takamiya M, Masuda S, Funahashi T, Tanaka T, Kudoh Y, Yamazaki J, Kushida N, Oguchi A, Kikuchi H, *et al.,* 1999. Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1. DNA Res. 6: 83-101.

Kawarabayasi Y, Sawada M, Horikawa H, Haikawa Y, Hino Y, Yamamoto S, Sekine M, Baba S, Kosugi H, Hosoyama A, Nagai Y, Sakai M, Ogura K, Otsuka R, Nakazawa H, Takamiya M, Ohfuku Y, Funahashi T, Tanaka T, Kudoh Y, Yamazaki J, Kushida N, Oguchi A, Aoki K, Kikuchi H. 1998. Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. DNA Res. 5: 55-76.

Klenk HP, Clayton RA, Tomb JF, White O, Nelson KE, Ketchum KA, Dodson RJ, Gwinn M, Hickey EK, Peterson JD, Richardson DL, Kerlavage AR, Graham DE, Kyrpides NC, Fleischmann RD, Quackenbush J, Lee NH, Sutton GG, Gill S, Kirkness EF, Dougherty BA, McKenney K, Adams MD, Loftus B, Venter JC, *et al.,* 1997. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. Nature. 390: 364-370.

Koebnik R. 1995. Proposal for a peptidoglycan-associating alpha helical motif in the C-terminal region of some bacterial cell -surface proteins. Mol. Microbiol. 16: 1269-1270.

Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, Bertero MG, Bessieres P, Bolotin A, Borchert S, Borriss R, Boursier L, Brans A, Braun M, Brignell SC, Bron S, Brouillet S, Bruschi CV, CaldwellB, Capuano V, Carter NM, Choi SK, Codani JJ, Connerton IF, Danchin A, *et al.,* 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature. 390: 249-156.

Larsen R., Thomas M. and Postle K. 1999. Proton motive force, ExbB and ligand bound FepA drive conformational changes in TonB. Mol. Microbiol. 31: 1809-1824.

Larsen R., Thomas M., Wood G. and Postle K. 1994. Partial suppression of the *E. coli* TonB transmembrane domain mutation D17V. by a missense mutation in ExbB. Mol. Microbiol. 13: 627-640.

Lazdunski C., Bouveret E., Rigal A., Journet L., Lloubès R. and Bénédetti H. 1998. Colicin import into *E. coli* cells. J. Bacteriol. 180: 4993-5002.

Lazzaroni J-C. and Protallier R. 1992. The *excC* gene of *E. coli* K-12 required for cell envelope integrity encodes the peptidoglycan associated lipoprotein Pal. Mol. Microbiol. 6: 735-742.

Lazzaroni J-C., Vianney A., Popot J-L., Bénédetti H., Samatey F., Lazdunski C., Portalier R. and Geli V. 1995. Transmembrane α-helix interactions are required for the functional assembly of the *E. coli* Tol complex. J. Mol. Biol. 246: 1-7.

Levengood S., Beyer W. and Webster R. 1991. TolA, a membrane protein involved in colicin uptake contains an extended helical region. Proc. Nat. Acad. Sci. USA. 88: 5939-5943.

Lubkowski J. Hennecke F., Pluckthun A. and Wlodawer A. 1999. Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. Structure Fold Des. 7: 711-722.

Müller M. and Webster R. 1997. Characterization of the *tol-pal* and *cyd* region of *Escherichia coli* K-12 transcript analysis and identification of two new proteins encoded by the *cyd* operon. J. Bacteriol. 179: 2077-2080.

Müller M., Vianney A., Lazzaroni J-C., Webster R and Portallier R. 1993. Membrane topology of the *E. coli* TolR protein required for cell envelope integrity. J. Bacteriol. 175: 6059-6061.

Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, McDonald L, Utterback TR, Malek JA, Linher KD, Garrett MM, Stewart AM, Cotton MD, Pratt MS, Phillips CA, Richardson D, Heidelberg J, Sutton GG, Fleischmann RD, Eisen JA, Fraser CM, *et al.,* 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature. 399: 323-329.

Parkhill J, Achtman M, James KD, Bentley SD, Churcher C, Klee SR, Morelli G, Basham D, Brown D, Chillingworth T, Davies RM, Davis P, Devlin K, Feltwell T, Hamlin N, Holroyd S, Jagels K, Leather S, Moule S, Mungall K, Quail MA, Rajandream MA, Rutherford KM, Simmonds M, Skelton J, Whitehead S. 2000. Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. Nature. 404: 502-506.

Parkhill J, Wren BW, Mungall K, Ketley JM, Churcher C, Basham D, Chillingworth T, Davies RM, Feltwell T, Holroyd S, Jagels K, Karlyshev AV, Moule S, Pallen MJ, Penn CW, Quail MA, Rajandream MA, Rutherford KM, van Vliet AH, Whitehead S, Barrell BG. 2000. The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. Nature. 403: 665-668.

Ray M.C., Germon P., Vianney A., Portalier R. and Lazzaroni J-C. 2000. Identification by genetic supression of Escherichia coli TolB residues important for TolB-Pal interaction. J. Bacteriol. 182: 821-824.

Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, Hickey EK, Peterson J, Utterback T, Berry K, Bass S, Linher K, Weidman J, Khouri H, Craven B, Bowman C, Dodson R, Gwinn M, Nelson W, DeBoy R, Kolonay J, McClarty G, Salzberg SL, Eisen J, Fraser CM. 2000. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. Nucl. Acids Res. 28: 1397-406.

Reichmann L. and Holliger P. 1997. The C-terminal domain of TolA is the coreceptor for filamentous phage infection of *E. coli*. Cell. 90: 351-360.

Rigal A., Bouveret E., Lloubès R, Lazdunski C. and Bénédetti H. 1997. The TolB protein interacts with the porins of *Escherichia coli*. J. Bacteriol. 179: 7274-7279.

Smith DR, Doucette-Stamm LA, Deloughery C, Lee H, Dubois J, Aldredge T, Bashirzadeh R, Blakely D, Cook R, Gilbert K, Harrison D, Hoang L, Keagle P, Lumm W, Pothier B, Qiu D, Spadafora R, Vicaire R, Wang Y, Wierzbowski J, Gibson R, Jiwani N, Caruso A, Bush D, Reeve JN, *et al.,* 1997. Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. J. Bacteriol. 179: 7135-7155.

Stanley P., Koronakis V. and Hughs C. 1998. Acylation of *Escherichia coli* Haemolysin: A unique protein lipidation mechanism underlying toxin function. Microbiol. Mol. Biol. Rev. 62: 309-333.

Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L, Mitchell W, Olinger L, Tatusov RL, Zhao Q, Koonin EV, Davis RW. 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science. 282: 754-9.

Sun T-P. and Webster R. E. 1987. Nucleotide sequencing of a gene cluster involved in entry of E colicins and single stranded DNA of infecting filamentous bacteriophages into *Escherichia coli*. J. Bacteriol. 169: 2667-2674.

Tettelin H, Saunders NJ, Heidelberg J, Jeffries AC, Nelson KE, Eisen JA, Ketchum KA, Hood DW, Peden JF, Dodson RJ, Nelson WC, Gwinn ML, DeBoy R, Peterson JD, Hickey EK, Haft DH, Salzberg SL, White O, Fleischmann RD, Dougherty BA, Mason T, Ciecko A, Parksey DS, Blair E, Cittone H, Clark EB, Cotton MD, Utterback TR, Khouri H, Qin H, Vamathevan J, Gill J, Scarlato V, Masignani V, Pizza M, Grandi G, Sun L, Smith HO, Fraser CM, Moxon ER, Rappuoli R, Venter JC. 2000. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. Science. 287: 1809-1815.

Thompson J.D., Higgins D.G., Gibson T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucl. Acids Res. 22: 4673-4680.

Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk HP, Gill S, Dougherty BA, Nelson K, Quackenbush J, Zhou L, Kirkness EF, Peterson S, Loftus B, Richardson D, Dodson R, Khalak HG, Glodek A, McKenney K, Fitzegerald LM, Lee N, Adams MD, Venter JC, *et al.,* 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature. 388: 539-547.

Vianney A., Lewin T., Beyer W. Lazzaroni J-C., Portallier R. and Webster R. 1994. Membrane topology and mutational analysis of the TolQ protein of E. coli required for the uptake of macromolecules and cell envelope integrity. J. Bacteriol. 176: 822-829.

Vianney A., Müller M., Clavel T., Lazzaroni J-C., Portalier R. and Webster R. 1996. Characterisation of the *tol-pal* region of *E. coli* K-12: translational control of *tolR* expression by *tolQ* and identification of a new open reading frame downstream of *pal* encoding a periplasmic protein. J. Bacteriol. 178: 4031-4038.

Webster R. 1991. The *tol* gene products and the import of macromolecules into *E. coli*. Mol. Microbiol. 5: 1005-1011.

White O, Eisen JA, Heidelberg JF, Hickey EK, Peterson JD, Dodson RJ, Haft DH, Gwinn ML, Nelson WC, Richardson DL, Moffat KS, Qin H, Jiang L, Pamphile W, Crosby M, Shen M, Vamathevan JJ, Lam P, McDonald L, Utterback T, Zalewski C, Makarova KS, Aravind L, Daly MJ, Fraser CM, *et al.,* 1999. Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. Science. 286: 1571-1577.

Zhou L., Srisatjaluk R., Justus D. and Doyle R. 1998. On the origin of membrane vesicles in gram negative bacteria. FEMS Microbiol. Lett. 163: 223-228.