The tadpole effect in Escherichia coli B/r

Binary fission of bacilliform bacteria produces two daughter cells that are inherently different — apart from any small variations in size and fluctuations in content that may arise, their outer (non-contiguous) poles are never the same age. This phenomenon has been investigated in the context of chromosome segregation (4), zonal growth (5) and, especially, asymmetric growth (2). In this last study, it was concluded that length and growth rate differ systematically according to the relative position of the cells in the family tree: the sister with the younger pole is shorter at inception but elongates more rapidly during constriction. There was no attempt to measure cell diameter as it was well beyond the resolution of the optics used.

In Escherichia coli B/r, there is a mechanism (1) that ensures that daughter cells are very nearly equal in length (7). Nonetheless, it has been observed under the electron microscope (8) that the thicker daughter tends to be shorter, imparting a tadpole-like form to the deeply-constricted mother cell. We have decided to look into this phenomenon to see whether it can be substantiated statistically.

As a source of raw data we took the very large sample, over 55000 cells, measured by Vardi (9) and described in detail by Grover & Woldringh (3). Prospective daughter cells were derived from deeply constricted mothers using the same criterion as before, yielding 637 mothers and 1274 daughters.

A certain caution must be exercised when we come to define the variables to be tested for correlation. Consider $D$, defined by $D = 2(D_1 - D_2)/(D_1 + D_2)$, where $D_1$ is the width of the first daughter cell and $D_2$ that of the second. It represents the difference in the widths of the daughters normalized by the mean width of their mother. But which is the ‘first’ daughter and which is the ‘second’? Their location in the data file is accidental, historical in nature and arbitrary, but not necessarily random. In fact, there are 306 cases with $D$ negative, 192 with $D$ zero, but only 139 with $D$ positive. Better to ask: ‘is the thicker sister also shorter?’ This is easily done by sequencing the daughters so that $D_1 > D_2$. We can then define $L$ to be $(L_1 - L_2)/(L_1 + L_2)$; the length of the thick daughter less the length of the thin, normalized to the length of their mother. The Pearson product-moment correlation coefficient $r$ between $D$ and $L$ is $-0.354$, which is very highly significant ($P < 0.001$); thicker sisters are indeed shorter. (Note that this and subsequent tests are carried out with $n$, the number of cases, set to the number of mothers, since siblinging information is fully redundant: the $D$ of one sister is the negative of the $D$ of the other, and similarly for $L$.)

There is still a problem with these results, however. Consider the case of a mother of uniform width that divides symmetrically; that is, $D = 0$ and $L = 0$. We would not expect, and certainly not want, such a cell to contribute to the width–length correlation. But the way the computation stands now, it does, as can be seen as follows. We ordered the prospective daughter cells so that there are no negative $D$ values; their average, $\bar{D}$, is therefore positive. Since the width–length correlation is negative, we would expect the average length $\bar{L}$ to be negative as well, and it is. Thus a mother with $D = 0$ and $L = 0$ has a lower-than-average $D$ value but a higher-than-average $L$ value and so makes the negative $D$–$L$ correlation even more negative.

The easiest way to remove this unwanted contribution is to drop all such cells from the calculation. Just to make sure, only cells with $D_1 - D_2 > 0.01$ $\mu$m were retained. The reduced sample now consists of 203 mothers, and the correlation coefficient is $-0.271$, weaker than before but still very highly significant ($P < 0.001$). We conclude that the thicker a daughter cell is compared to her sister, the shorter it is as well, statistically speaking.

A negative width–length correlation would seem to imply that the relative dispersion in cell volume is less than in cell length. Woldringh and I have recently published an article (3) in which we show that length rather than volume is the controlling dimension at such cell-cycle events as cell division. Among the evidence we present in support of that thesis is data from the same set as used here showing that the coefficient of variation (standard deviation – mean) of cell volume at division is about twice that of cell length. Is there a contradiction? Not necessarily.

The apparent contradiction disappears once we realize that we are dealing with different dispersions: the one here involves the distribution of sister lengths about their
Escherichia coli's uropathogenic-specific protein: a bacteriocin promoting infectivity

In 2000, Kurazono et al. (3) reported the fortuitous discovery of a gene encoding a protein, designated Usp, from an uropathogenic Escherichia coli (UPEC) strain. E. coli strain Z42 was isolated from a patient suffering from an inflammation of the prostate gland (prostatitis). The DNA element carrying the usp gene was found on the chromosome of a large number of UPEC strains (4) and is significantly more frequently associated with UPEC strains than with faecal E. coli strains from healthy individuals (8). In addition, the contribution to virulence of the Usp protein in uropathogenic E. coli was assessed using an experimental mouse model. It was observed that an E. coli strain possessing the usp gene enhanced the infectivity of E. coli, supporting the hypothesis that usp is located on a small (42 kb) pathogenicity island (PI) (8). Downstream of usp three small ORFs (orfU1, orfU2 and orfU3) were identified putatively encoding proteins of 98, 97 and 96 aa, respectively. No function could be attributed to orfU1, orfU2 and orfU3 from the experimental mouse model. At present, the sequences of seven PIs from representative uropathogenic E. coli strains are available in the databases (4). All PIs contain an usp gene preceding one, two or three of the smaller ORFs (orfU1, orfU2 and orfU3).

We stumbled on the Usp protein in the course of homology searches for relatives of the S-type pyocins. Pyocins, produced by Pseudomonas aeruginosa strains, belong to a large and heterogeneous group of proteins known as bacteriocins. These are ribosomally synthesized antimicrobial molecules produced by bacteria and directed against closely related bacteria (6). S-type pyocins are chromosomally encoded bacteriocins sharing functional and structural characteristics with a subgroup of E. coli bacteriocins (nuclease colicins). S pyocins and nuclease colicins are both produced as protein complexes consisting of a killing protein with nuclease activity and an immunity protein. The immunity protein, encoded by an innm gene closely linked to the bacteriocin gene, provides protection against the killing activity inside the producer cell. In S pyocins, the N-terminal region, which is involved in receptor recognition, is followed by a translocation domain and a Dnase/immunity protein-binding domain. In nuclease colicins, the order of receptor recognition and translocation domains is reversed. While investigating the distribution of S pyocin-like genes throughout the γ-Proteobacteria, we frequently encountered Usp proteins showing significant homology to our input sequences, although Usp proteins are claimed to have no similarity with any proteins in the databases (3, 4). BLASTp searches (http://www.ncbi.nlm.nih.gov/BLAST/) with the Usp proteins as query sequences pinpoint S pyocins S1, S2 and AP41 as well as colicin E9 as the closest relatives. These bacteriocins are all known to possess endonuclease activity. Pairwise alignments of the seven known Usp protein sequences with endonuclease bacteriocins revealed a C-terminal domain (138 aa) with 40–45% identity to the Dnase-like domain of the S pyocin AP41 (Fig. 1a). In addition, we detected an HNH endonuclease motif (PF01844) in the C-terminal regions of all seven Usp proteins (Fig. 1a). Such a conserved motif is present in the C-terminal domain of Dnase-type colicins and S pyocins (5). We investigated the degree of homology between the putative Dnase domains of the Usp proteins. Results showed that Usp proteins can be divided into two groups: E. coli strains Z42, C72, Z25 and E25 encode highly homologous Usp proteins (UspI group) as do strains Z16, Z13 and P17 (UspII group). Sequence variation is located exclusively in the C-terminal region of the proteins (Fig. 1a). In front of each of the putative Dnase domains, we detected a possible translocation domain showing approximately 35% similarity with part of the corresponding domains of S pyocins (Fig. 1b). The N-terminal region of Usp proteins exhibits no homology to any known protein. Usp protein sequences are claimed to include a 24 aa putative signal sequence (3), although we could not confirm this with a SignalP search (http://www.cbs.dtu.dk/services/SignalP-2.0/). However, careful inspection of the upstream region of usp genes revealed an N-terminal extension with a conserved motif found in Hcp proteins (PB022986) (Fig. 1a, b). Hcp of Vibrio cholerae has been characterized as a secreted haemolysin-coregulated protein of unknown function, dispensable for V. cholerae infection in an infant mouse model (7). Genes for Hcp-like proteins are found in the genomes of several human pathogens, such as the enterohemorrhagic E. coli strain O157:H7, P. aeruginosa, Vibrio parahaemolyticus, and Yersinia pestis. An Hcp-like protein was detected in culture supernatants of the phyo-

Fig. 1. See facing page. (a) Alignment of predicted amino acid sequences of the Usp proteins of Escherichia coli Z42 (UspL.Ec, deduced from gi.18147638) and Z16 (UspII.Ec, deduced from the incomplete sequence described by gi.18147657). The N-terminal region of UspI and UspII is aligned with Hcp-like proteins of E. coli (Hcp.Ec, gi.2906844), Pseudomonas aeruginosa (Hcp.Pa, gi.1559846) and Vibrio cholerae (Hcp.Vc, gi.1488371). The C-terminal region of Usp and UspII shows homology to the nuclease family of colicins and pyocins. Representative members of this family are colicin E7 (ColE7.Ec, gi.12644448) and colicin E9 (ColE9.Ec, gi.1705745) of E. coli, pyocin S1 (ProS1.Pa, gi.286180) and pyocin

Microbiology 148, June 2002

1604
AP41 (PyoAP41_Pa, gi:216904) of *P. aeruginosa* and klebicin B (KlebB, Kp, gi:11181544) produced by *Klebsiella pneumoniae*. Conserved Pfam domains are depicted with boxes and numbered accordingly (http://www.sanger.ac.uk/Software/Pfam). (b) Organization of the functional domains inferred for the predicted UspI and putative immunity protein sequences (OrfU1, OrfU2 and OrfU3) of *E. coli* strain Z42. Block arrows indicate the transcription orientation of the genes. Domain structure is represented by different shading patterns (see inset). Identical shading implies sequence homology, except for the translocation and receptor recognition domains of colicin E9. Unshaded regions show no significant homology to any proteins in the databases. Apparent bacteriocin gene remnants were detected upstream of *orfU2* and *orfU3*. A similar gene organization was observed for the six additional Usp-encoding genomic sequences from *E. coli* strains Z16 (gi:28147650), Z25 (gi:18147646) and Z25 (gi:18147646). An asterisk denotes a potential frameshift. (c) Molecular phylogeny of the putative Usp immunity proteins (OrfU1, OrfU2 and OrfU3) and representative immunity proteins of nuclease bacteriocins. The alignment used for the phylogenetic analysis was generated using the ClustalW algorithm. The tree topology was inferred using the PHYLIP software package with 1000 replicates (http://evolution.genetics.washington.edu/phylip.html). This tree is the best tree by parsimony. A tree derived from distance analysis with a neighbour-joining algorithm showed identical topology. Bootstrap values (parsimony/distance) are indicated on the appropriate branches. ImmPyoAP41 (gi:216905) and ImmPyoS1 (gi:286180) are the corresponding immunity proteins for pyocin AP41 and pyocin S1 produced by *Pseudomonas aeruginosa*. ImmKlebB (gi:5070707) is the immunity protein associated with klebicin B, both encoded on the *Klebsiella pneumoniae* plasmid pKlebB-K17/80. ImmColE7 (gi:1942132) and ImmColE9 (gi:124403) are immunity proteins of cognate colicins E7 and E9 encoded by Col plasmids pColE7-317 and pColE9-J of *E. coli*, respectively.
pathogen *Pseudomonas syringae* pv. *rubicola* (2). We anticipate the actual length of Usp proteins to be approximately 600 aa (instead of 396 aa) including the Hcp-like N-terminal extension possibly involved in Usp secretion.

The predicted proteins encoded by *orfU1*, *orfU2* and *orfU3*, contain a conserved motif characteristic for immunity proteins of the nuclease bacteriocins (PF01320). Significant levels of homology (40–50% identity) are observed when comparing *OrfU1*, *OrfU2* and *OrfU3* with the immunity proteins of the corresponding DNase-type pyocins and colicins. Investigation of the organization of the different PIs shows *orfU1* to be closely linked with *uspI*, whilst *orfU2* is located immediately downstream of genes encoding UspII proteins. This observation may be interpreted as if *orfU1* would be the *imm* gene for UPEC strains producing Uspl bacteriocin, while *orfU2* would render immunity towards UspII. A phylogenetic tree constructed from a protein alignment of the putative immunity proteins encoded by the pathogenicity islands of the UPEC strains as well as representative nuclease-type pyocins and colicins. Investigation of the organization of the different PIs shows *orfU1*, *OrfU2* and *OrfU3* proteins form three distinct clusters. Within each of these clusters, OrfU proteins of UPEC strains producing Uspl versus UspII also clearly separate, reflecting the relatedness of the corresponding Usp proteins.

While colicins are typically encoded on Col plasmids, Usp may represent a novel type of bacteriocin. Competition of colicinogenic and sensitive *E. coli* in the rat urinary tract has been described (1, 6). We hypothesize that the Usp protein acts as a bacteriocin against competing *E. coli* strains occupying the same niche, thereby enhancing their infectivity in the urinary tract environment. *E. coli* armed with bacteriocin genes can protect itself against the cytotoxic activity of its own bacteriocin by co-synthesizing the cognate immunity protein. The additional immunity genes downstream of the linked immunity gene could render the producing cell resistant towards DNase bacteriocins produced by its competitors. Such multiple immunity systems have been described for *E. coli* and other γ-Proteobacteria (5). Further characterization of the Usp protein should include bacteriocin assays to identify sensitive *E. coli* strains and biochemical analysis of the predicted endonuclease activity. Confirmation of the bacteriocin nature of Usp, would not only identify a novel type of bacteriocin of *E. coli* but would also provide evidence for a role of the bacterial toxin in host infectivity, based on increased competitiveness.

**Acknowledgements**

We acknowledge the Federal Office for Scientific, Technical and Cultural Affairs (DWTC) for support through Grant IUAP-V-03.

Annabel H. A. Parret and René De Mot

Centre of Microbial and Plant Genetics, Katholieke Universiteit Leuven, Kasteelpark Arenberg 20, 3001 Heverlee, Belgium.

Author for correspondence: Annabel H. A. Parret. Tel: +32 16 329657. Fax: +32 16 321966. e-mail: annabel.parret@agr.kuleuven.ac.be