Evidence for the presence in \textit{Burkholderia pseudomallei} of a type III secretion system-associated gene cluster

C. Winstanley, B. A. Hales and C. A. Hart*

Department of Biomedical Sciences, University of Bradford, Bradford, West Yorkshire BD7 1DP and *Department of Medical Microbiology and Genito-Urinary Medicine, University of Liverpool, PO Box 147, Liverpool L69 3BX

\textit{Burkholderia pseudomallei}, the causative agent of melioidosis, contains a cluster of putative genes homologous to those encoding HpaP, HrcQ, HrcR, HrcS and HrpV in the plant pathogen \textit{Ralstonia solanacearum}. In \textit{R. solanacearum}, these genes form part of a type III secretion-associated pathogenicity island. The order of the genes in \textit{B. pseudomallei} is directly equivalent to that found in \textit{R. solanacearum}. The \textit{B. pseudomallei} proteins share 49.5\% (HpaP), 52.6\% (HrcQ), 80.0\% (HrcR), 72.1\% (HrcS) and 46.7\% (HrpV) similarity, respectively, with their equivalent \textit{R. solanacearum} proteins. The presence of type III secretion-associated genes in \textit{B. pseudomallei} pathogens suggests a possible role for type III secretion systems in the pathogenicity of this organism.

Introduction

\textit{Burkholderia} (formerly \textit{Pseudomonas}) \textit{pseudomallei} is the causative agent of melioidosis, an often fatal infection endemic in areas of South East Asia and Australia, where it is a major cause of community-acquired septicaemia [1]. Although some potential virulence factors have been identified, overall knowledge of the pathogenicity of \textit{B. pseudomallei} is limited. However, a number of its secreted products, including protease, haemolysin, lipase and lecithinase [2, 3] have been linked to virulence.

Pathogenicity islands (PIS), virulence genes found in large contiguous groups, have been identified in several bacterial pathogens [4, 5]. In some gram-negative bacteria, PIS have been associated with type III secretion systems [4, 6, 7], which are made up of a number of proteins homologous with components of flagellum-specific export apparatus, and are involved in delivering virulence factors directly to host cells [4, 6]. In contrast to type I and type II secretion systems [8, 9], type III secretion is triggered by a pathogen coming into close contact with host cells [4, 6, 10].

Type III secretion system-associated PIS have also been characterised in a number of plant pathogens [6, 11, 12], including \textit{Ralstonia} (formerly \textit{Burkholderia}) \textit{solanacearum}, the causative agent of bacterial wilt in various crops and plants [13]. In this organism, five transcription units of the \textit{hrp} gene cluster (involved in development of the hypersensitive response, HR) are required for the secretion of the HR-inducing protein, PopA1. The nucleotide sequence for the entire region (18,300 bp) of the \textit{R. solanacearum} strain GMI1000 \textit{hrp} gene cluster implicated in the secretion of PopA1 has been reported, including all five transcription units, encoding a total of 20 Hrp proteins [13].

This study reports the use of a \textit{R. solanacearum} \textit{hrp} gene probe to identify, clone and sequence \textit{B. pseudomallei} genomic DNA with homology to type III secretion-associated genes.

Materials and methods

\textit{Bacterial strains and plasmids}

\textit{B. pseudomallei} E503 was isolated from a patient with melioidosis in Malaysia [14]. \textit{R. solanacearum} GMI1000 and cosmid pVir2, which carries most of the \textit{hrp} gene
cluster of strain GM1000 [15], were provided by Dr M. Arlat and P. Barberis (Laboratoire de Biologie Moléculaire des Relations Plantes-Microorganismes, INRA-CNRS, BP27, 31326 Castanet Tolosan Cedex, France).

Construction of gene libraries

Genomic DNA, extracted from B. pseudomallei E503 as described previously [16], was employed to construct gene libraries with the SuperCos 1 Cosmid Vector Kit (Stratagene) and the conditions recommended by the supplier.

Identification of cosmid clones homologous to hrp probes

Oligonucleotide primers, obtained from Genosys Biotechnologies (Europe), Cambridge, for PCR amplification were designed with the sequence information available for the hrp locus of R. solanacearum [13]. A combination of the primers BS7 (5'-TGTCAGCGATAAGTGTTGACCG-3') and BS8 (5'-ATTGCACTCAGGCAAGCACCA-3') was employed to generate an amplified product of 1010 bp (location 3106–4116 GenBank Accession no. z14056) encompassing the whole of hrcR (hrpT) and part of hrcQ (hrpQ). pVir2 cosmid DNA was isolated with a Midi-preparation kit (Qiagen); 1 µl of various dilutions of pVir2 DNA was used directly in 25-µl volumes containing 2 units of Dynazyme (Flowgen), 200 nM of each primer (BS7 and BS8), 1 × Dynazyme buffer and 100 µM nucleotides (dATP, dCTP, dGTP, dTTP). Amplifications were carried out in a MiniCycler (Genetic Research Instrumentation) for 30 cycles consisting of 95°C (1 min), 60°C (1 min) and 72°C (2 min) with an additional extension time at 72°C (10 min) after completion of the 30 cycles. At the end of the amplification, 10-µl samples were subjected to electrophoresis on a standard agarose 1.0 w/v) gel electrophoresis by standard procedures. The 1010-bp amplified product was labelled with digoxigenin-11-2'-dUTP (DIG; Boehringer Mannheim) by repeating the PCR reaction with 60 µM DIG in a total reaction volume of 50 µl. The labelled product was used as a probe to identify homologous gene-containing clones from a B. pseudomallei E503 gene library. The presence of DIG on colony blots was detected by anti-DIG-AP Fab fragments and the chemiluminescent substrate CDP-Star (Boehringer Mannheim) in the procedure recommended by the supplier.

Subcloning of hybridising genomic regions

Cosmid DNA was isolated with a Midi-preparation kit (Qiagen) and subjected to restriction endonuclease (Helena Biosciences) digestion and agarose (0.7% w/v) gel electrophoresis by standard procedures. Smaller hybridising fragments, identified by Southern blot analysis of digested cosmid clones, were subcloned into the plasmid vector pUC19 (Helena Biosciences).

Nucleotide sequencing and computer analyses

DNA was purified from putative clones with a QIAprep Spin Miniprep Kit (Qiagen). Both strands of the cloned insert DNA were sequenced by primer walking at the University of Liverpool DNA Sequencing Service with vector and internal oligonucleotide primers.

Nucleotide sequence alignments, G + C mol% values, determination of amino acid composition, predicted protein mass and alignments of predicted proteins with other related proteins (retrieved from EMBL, GenBank, PIR or SwissProt [17]) were performed with GAP, PILEUP, COMPOSITION, PEPTIDESORT and FASTA programs from the GCG sequence analysis software package (Genetics Computer Group, University of Wisconsin). Hydrophobicity profiles and predictions of membrane-spanning regions were obtained with the PEPTIDESTRUCTURE and PEPPLOT programs, employing the hydropathy measure of Kyte and Doolittle [18] and the measure of helical hydrophobic moment [19].

Accession number

The 4490-bp B. pseudomallei nucleotide sequence reported in this study has been lodged under the GenBank accession number AFO42488.

Results

Identification of cosmid clones containing putative type III secretion genes

With a DIG-labelled PCR-amplified product as a probe, it was possible to identify B. pseudomallei E503 cosmids showing strong homology to the region of the R. solanacearum GM1000 genome encoding HrcR and part of HrcQ. To locate the hybridising region more accurately, DNA was isolated from one representative hybridising cosmid clone, digested with several restriction endonucleases (including SalI and PstI) and probed by Southern blot hybridisation.

Nucleic acid and predicted protein sequence comparisons

Hybridising fragments of B. pseudomallei E503 genomic DNA of 2.6 kb (generated with SalI) and c. 3.0 kb (generated with PstI) were subcloned into pUC19 and subjected to DNA sequencing (Fig. 1). The overlapping fragments enabled a nucleotide sequence of 4490 bp to be determined. A database comparison indicated that by far the best homology for this sequence was to regions of the R. solanacearum type III secretion-associated PI [13] (GenBank Accession no. z14056). Within the B. pseudomallei nucleotide sequence, five complete open reading frames (ORFs) were identified. Predicted protein sequences, obtained for each of the five ORFs, suggested the presence of putative genes homologous to those encoding HpaP, HrcQ (HrpQ), HrcR (HrpT), HrcS (HrpU) and HrpV genes of the R. solanacearum type III...
secretion system, and that the genes were clustered in the same order [13]. The Hrc designations are as suggested by Alfano and Collmer [12]. Previous designations, as reported for *R. solanacearum* by Van Gijsegem et al. [13], are indicated in brackets. Under the unified nomenclature suggested by Hueck [6], three of the *B. pseudomallei* predicted proteins have been designated SctQ (HrcQ homologue), SctR (HrcR homologue) and SctS (HrcS homologue). Because they lack homologues in the majority of type III secretion systems reported to date, the HpaP (ORF1) and HrpV (ORF2) homologues are not included in the unified nomenclature. SctR and SctS had the highest homologies with previously reported proteins. SctR/SctS homologues are transmembrane protein components of type III secretion systems, including YscR/YscS of *Yersinia pestis*, Spa24/Spa9 of *Shigella flexneri* and SpaP/SpaQ of *Salmonella enterica* serotype Typhimurium, all from genes important in animal pathogenicity [4]. Alignments of SctR/SctS with equivalent proteins from a number of type III secretion systems are shown in Fig. 2. Hydrophobicity profiles of *B. pseudomallei* SctR/SctS and *R. solanacearum* HrcR/HrcS were used to identify a number of transmembrane-predicted regions, which were in broad agreement with the previous report of Van Gijsegem et al. [13]. In the SctR homologues, all but one of the four predicted transmembrane regions were in highly conserved regions of the protein (Fig. 2a, Table 1). The SctS homologues contained two predicted transmembrane regions separated by a short hydrophilic region (Fig. 2b, Table 1).

The predicted protein sequences for the other three putative *B. pseudomallei* proteins were aligned with their *R. solanacearum* homologues (Fig. 3). Predicted features and sequence similarity/identity measurements for all the proteins are presented in Table 1. Comparison of the *B. pseudomallei* HpaP homologue sequence against protein databases confirmed that by far the best homology was with *R. solanacearum* HpaP. The HpaP protein of *R. solanacearum* is unusually rich in proline (15% of total amino acid residues), a feature that may be important in tertiary structure [20]. The *B. pseudomallei* HpaP homologue contained 22 proline residues (9.7%), making it the third most frequent residue. SctQ homologues exhibit homology only at the C-terminal region [13], making comparisons of the whole protein difficult. Alignment of the last 80 C-terminal residues of *B. pseudomallei* SctQ and *R. solanacearum* HrcQ reveals 72.5% similarity and 42.5% identity (Fig. 3). The putative HrpV homologue is the least conserved of the *B. pseudomallei* predicted protein sequences, but its hydrophobicity profile, pI and Mr are similar to *R.
Fig. 2. Comparison of putative *B. pseudomallei* SetR and SetS protein sequences with homologues from other organisms encoding type III secretion systems. Alignment of *B. pseudomallei* (BURK) sequences with related proteins from *Salmonella* (SALM), *Shigella* (SHIG), *Pseudomonas* (PSEU), *Erwinia* (ERWI), *Ralstonia* (RALS), *Xanthomonas* (XANT), *Yersinia* (YERS), *Rhizobium* (RHIZ) and *Escherichia* (ESCH) are shown for SetR (a) and SetS (b).

Residues conserved in >50% of the sequences used are shaded. Thick black lines indicate the location of hydrophobic regions.

*solanacearum* HrpV (Table 1), indicating the possibility of similar roles in both organisms. Although a complete ORF is not yet available, there is evidence in *B. pseudomallei* for nucleotide sequence homology with the *R. solanacearum* gene encoding HrcV (HrpO) (65.8% identity in 260 bp) upstream of the hpaP-hrcQ-hrcR-hrcS-hrpV region. When translated, a downstream incomplete ORF yields a peptide sequence of 93 residues with 56.7% similarity (GAP alignment) to the 97 N-terminal residues of *R. solanacearum* HrpW.

**Discussion**

PISs are thought to offer bacteria the possibility of acquiring, in one genetic event, a virulence trait, such as secretion of specific toxins in response to host cell
proximity. The presence in *B. pseudomallei* of adjacent putative SctR and SctS coding regions, exhibiting high homology to equivalent *R. solanacearum* proteins, and three other ORFs with predicted proteins exhibiting homology to equivalent *R. solanacearum* type III secretion components, suggests a cluster of related genes similar to those located in the PI of *R. solanacearum*. The gene organisation for *B. pseudomallei* and *R. solanacearum* is identical, indicating that similar PI structures may reside in these two organisms, despite the fact that one is pathogenic for man and the other for plants. The order of the sctQRS gene cluster is also conserved in *Y. pestis* [21], *Sh. flexneri* [22], *S. enterica* [23] and the plant pathogens *Erwinia amylovora*, *Pseudomonas syringae* and *Xanthomonas campestris* [12]. Both SctR and SctS homologues contain predicted membrane-spanning regions and are thought to reside in the inner membrane [6].

Proteins homologous to *R. solanacearum* HpaP and HrpV have not been reported previously in animal pathogens, although an HpaP homologue has been reported in *X. campestris*, a plant pathogen that has considerable similarities to *R. solanacearum* in the organisation of its type III secretion system genes [24]. The actual sequence of this homologue is not available for analysis. Although the homology between HpaP and its putative *B. pseudomallei* homologue is lower than that observed for SctR and SctS, the similarity in properties suggests that the role of the putative protein may be similar in both organisms. As yet, no function has been ascribed to this protein, which has no hydrophobic domains [20]. The *R. solanacearum* HrpV protein, which exhibits the lowest homology with its proposed *B. pseudomallei* equivalent, has also not been ascribed a function. Although the role of *R. solanacearum* HpaP and HrpV and their putative *B. pseudomallei* equivalents is not known, the observed divergence between the two organisms suggests that these two proteins are not confined by the same structural constraints as SctQ, SctR or SctS. It may be that HpaP and HrpV homologues have a role specific in adaptation to a host or secretion of different virulence factors. Sequence similarities for SctQ homologues are restricted to their 80 C-terminal amino acids [6]. It has been suggested that this protein may provide a link between common and species-specific components of type III secretion systems [6]. The presence, at either end of the gene cluster, of nucleotide sequences homologous to *R. solanacearum* *hrpC* and *hrpW* strongly suggests that a gene order is maintained between *B. pseudomallei* and *R. solanacearum* beyond the five putative genes reported in this paper. More sequencing will be required to determine fully the extent of the *B. pseudomallei* putative PI and its relationship to other type III secretion system-associated PIs.

It has been observed that whilst the majority of genes in individual type III secretion systems encode proteins with homologues in other systems, each respective system may contain genes not widely observed in other systems and often unique to itself [6]. In the type III secretion systems of *R. solanacearum*, *X. campestris* and *B. pseudomallei*, there is good conservation of the gene orders. This supports the notion of an ancestral type III secretion system acquired by one organism from another. The relatively high similarities across species barriers amongst the proteins supports this idea. However, neither the homology nor the G + C mol% content is uniform throughout the *B. pseudomallei* gene cluster. The overall G + C content for the 4490-bp sequence (63.1 mol%) is lower than the value obtained for the *R. solanacearum* *hrp* locus (18300 bp; 69 mol%), but is consistent with values reported for *Burkholderia* coding regions – including the *fitC* gene of *B. pseudomallei* (65 mol%) [14], the *fitC* gene (64 mol%) [25] or cable pili pilin gene (*cblA*) of *B. cepacia* (62 mol%) [26], albeit lower than the value reported for *B. cepacia* genomic DNA (67 mol%) [27]. However, there is fluctuation within the 4490-bp sequence. For example, the G + C value for the putative srcR coding region is 57 mol%. A lower than average G + C content is also found in the *hrcR* gene of *R. solanacearum* (64.4%). Fig. 1 demonstrates the variation in G + C content, and the correlation with conservation between *B. pseudomallei* and *R. solanacearum* throughout the gene cluster. In many organ-

### Table 1. Predicted properties of putative type III secretion-associated *B. pseudomallei* proteins

<table>
<thead>
<tr>
<th>Coding gene*</th>
<th>Similarity/</th>
<th>Identity(%)</th>
<th>Length</th>
<th>pI</th>
<th>M+</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>HpaP</td>
<td>49.5/28.4</td>
<td>227</td>
<td>6.8</td>
<td>24828</td>
<td>No strongly hydrophobic domains</td>
<td></td>
</tr>
<tr>
<td>hpaQ</td>
<td>52.6/27.6</td>
<td>311</td>
<td>5.6</td>
<td>33358</td>
<td>Aligns well only at the C-terminus; no strongly hydrophobic domains</td>
<td></td>
</tr>
<tr>
<td>HrpV</td>
<td>80.0/66.0</td>
<td>216</td>
<td>7.7</td>
<td>23662</td>
<td>A number of hydrophobic regions (resembling the equivalent <em>R. solanacearum</em> protein)</td>
<td></td>
</tr>
<tr>
<td>HrpV</td>
<td>72.1/55.8</td>
<td>87</td>
<td>5.5</td>
<td>9193</td>
<td>Two hydrophobic domains separated by a short hydrophilic region (as for the equivalent <em>R. solanacearum</em> protein)</td>
<td></td>
</tr>
<tr>
<td>HrpV</td>
<td>46.7/24.9</td>
<td>317</td>
<td>10.3</td>
<td>34417</td>
<td>No strongly hydrophobic domains</td>
<td></td>
</tr>
</tbody>
</table>

*Equivalent R. solanacearum genes are indicated in brackets. |

1Similarity and identity between *B. pseudomallei* and *R. solanacearum* homologues are indicated. |

2Values for *R. solanacearum* homologues are shown in brackets. |

### References

[21] Sh. flexneri [22], *S. enterica* [23] and the plant pathogens *Erwinia amylovora*, *Pseudomonas syringae* and *Xanthomonas campestris* [12]. Both SctR and SctS homologues contain predicted membrane-spanning regions and are thought to reside in the inner membrane [6].
Fig. 3. Comparison of putative *B. pseudomallei* proteins with *R. solanacearum* HrcQ, HpaP and HrpV. GAP alignments are shown for (a) the last 80 C-terminal residues of *B. pseudomallei* SctQ (upper) against the equivalent region of *R. solanacearum* HrcQ (lower); (b) *B. pseudomallei* predicted protein product of ORF1 (upper) against *R. solanacearum* HpaP (lower); (c) *B. pseudomallei* predicted protein product of ORF2 (upper) against *R. solanacearum* HrpV (lower). Identical residues are indicated by a vertical line. Two dots represent a greater similarity than one dot.
isms, the G + C mol% content of the type III secretion system genes has been reported to be lower than the mean genomic value, suggesting that the ancestral type III secretion system genes may have evolved in a low G + C mol% host [6]. Recently, a type III secretion system was identified in chlamydia [28], a potential candidate for such a role. In the B. pseudomallei gene cluster, the G + C mol% content is lower in those areas with the greatest degree of conservation (Fig. 1).

Although it is believed that type III secretion systems are acquired in intact genetic blocks, considerable degrees of variation in conservation of proteins have been observed [6]. There are a number of examples, in type III secretion systems, where genes are conserved in relative location whilst exhibiting low sequence homologies [6]. Whilst SetQ, SetR and SetS homologues are found in all type III secretion systems, HpaP and HprV homologues are restricted to a group including R. solanacearum, B. pseudomallei, and in the case of HpaP, X. campestris. Furthermore, the protein homologies exhibited by these proteins (not thought to be integral membrane-associated components of the type III secretion structure) are far lower than for SetQ (C-terminal region), SetR or SetS, despite their positional conservation. This variation is consistent with the idea that certain proteins have undergone greater evolutionary change than others, with possible roles specific to individual secreted proteins or adaptation to different hosts, and with the higher G + C mol% content, more typical of Burkholderia, observed in the less conserved regions.

In R. solanacearum, a gap between HpaP and HrcQ (HrpQ) coding regions defines the beginning of transcription unit 4, although the exact location of a promoter is not known [13]. In B. pseudomallei, the presence of a putative ribosome binding-site (AGGA) has been used to identify the probable start codon. Although no promoter sequence has been identified, the gap between the first and second ORFs in B. pseudomallei (Fig. 1) may be indicative of a similar gene organisation and regulation between the two organisms.

The role of a B. pseudomallei type III secretion system has yet to be elucidated, but the presence of such systems has been implicated in the pathogenicity of several other gram-negative bacterial pathogens [4, 6, 10]. One of the major goals in further characterising a putative B. pseudomallei type III secretion system is to identify potential secreted virulence factors. The type III secretion pathways of plant pathogenic bacteria, found to form two distinct groups, have been characterised in E. amylovorora, P syringae, X campestris and R. solanacearum. The Erwinia Hrp genes are closely related to the P. syringae Hrp genes (group I) whilst Ralstonia and Xanthomonas genes form a second group (group II) [12]. Group I and group II proteins are as distantly related to each other as they are to Yersinia type III secretion genes. The genes encoding the secreted proteins (harpins) are located adjacent to the secretion system genes in Erwinia (hpnN) and Ralstonia (popA), but within the secretion system gene cluster in P. syringae (hprz2). This proximity to the PI gene clusters, although not always the case, suggests that sequencing approaches could be used to identify a gene encoding a secreted product.

Michiels and Cornelis [29] demonstrated that successful secretion of the Yop proteins by the Yop type III secretion system of Yersinia was dependent on the N-terminal regions. Yet there was little homology between the N-terminal regions of the different secreted proteins. The authors suggested that the secretion signal may be conformational. More recently Anderson and Schneewind [30] have suggested that the signal may be encoded in the mRNA rather than the protein. This indicates that the gene encoding a secreted protein would not be readily identified by the use of genes encoding other secreted proteins as probes. However, Yahr et al. [31] reported some N-terminal homology between the different secreted proteins in P. aeruginosa. In addition, the authors were able to identify several proteins secreted by a type III secretion system by comparing extracellular protein production in wild-type and type III secretion system mutants, an approach which may also prove fruitful in B. pseudomallei now that the genes of a putative secretion system have been identified.

B. pseudomallei is a facultative intracellular pathogen [32]. Effector molecules secreted by type III secretion systems have been implicated in stimulation of entry into normal non-phagocytic cells by Shigella spp. [33] and S. Typhimurium [34], although the principal function of type III secretion is thought to be protein translocation into host cells [10]. Type III secretion systems have been widely identified as playing an essential role in interactions between bacterial pathogens and their hosts. Therefore, the presence of a type III secretion system gene cluster in B. pseudomallei is indicative of some role in the pathogenicity of this organism. The sequence data reported in this study will allow the production of specifically targeted mutants and demonstration of the importance of the gene cluster in pathogenicity by the use of animal models. Hardt and Galán [35] reported a type III-secreted salmonella protein with homology to an avirulence determinant of the plant pathogen X. campestris. The identification of B. pseudomallei secreted proteins will make it possible to determine any similar relationships between B. pseudomallei and R. solanacearum proteins.

It has been suggested that specific components of type III secretion systems could be targeted in order to inhibit the delivery of virulence factors to plant
pathogens [11]. Given the widespread emergence of studies indicating the importance of type III secretion systems in human bacterial pathogens, this intriguing possibility should not be discounted as a future therapeutic goal in man.

This work was supported by an award to C.W. from The Wellcome Trust (grant no. 044249/PMG/VW). We thank Dr Matthieu Arlat and Patrick Barberis (Laboratoire de Biologie Moleculaire des Relations Plantes-Microorganismes, INRA-CNRS, BP27, 31326 Castanet Tolosan Cedex, France) for kindly providing strains. We also thank Angela Rosin for DNA sequencing. This work benefited from the use of the SEQUENT facility, Daresbury.

References