Conjugal type IV macromolecular transfer systems of Gram-negative bacteria: organismal distribution, structural constraints and evolutionary conclusions

Thien B. Cao and Milton H. Saier, Jr

Author for correspondence: Milton H. Saier, Jr. Tel: +1 858 534 4084. Fax: +1 858 534 7108. e-mail: msaier@ucsd.edu

Department of Biology, University of California at San Diego, La Jolla, CA 92093-0116, USA

Keywords: conjugation, virulence, DNA transfer, protein secretion, Agrobacterium

Overview

Type IV secretory pathway (IVSP) systems are multi-component, transenvelope complexes in Gram-negative bacteria that translocate proteins and nucleoprotein complexes from donor cells to recipient cells in processes related to bacterial conjugation. At least ten protein constituents have been shown to mediate protein or nucleoprotein translocation, but the phylogenetic relationships of these proteins have not previously been defined. We have identified all recognizable homologues of the agrobacterial VirB2-11 proteins, retrieved their sequences from the databases, and characterized these homologues with respect to size, topology and organismal source. The homologous sequences were aligned in preparation for derivation of mean hydropathy, similarity and amphipathicity plots as well as phylogenetic trees. The results allowed us to make structural and functional predictions and show that although these systems have been repeatedly exchanged between Gram-negative bacteria, the constituents of these complex systems have not undergone appreciable shuffling during evolutionary divergence. This last observation further suggests that macromolecular transfer occurs by a concerted mechanism. Some homologues of IVSP constituents are likely to provide functions in some organisms that are unrelated to IVSP-mediated secretion. For example, homologues of VirB8, 9, 10 and 11 found in Helicobacter pylori, Campylobacter jejuni and Rickettsia prowazekii, as well as species of Wolbachia, may serve virulence-related functions not requiring a complete type IV secretion system. Both of the two IVSP-associated ATPases (VirB4 and VirB11 homologues) are found in prokaryotic organisms that lack the other constituents of the system, and VirB11 homologues are also found in archaea. It is proposed that IVSP systems evolved late as mosaic secretory systems that at least in part derived their constituents from other pre-existing sources, initially for conjugation with other prokaryotes and later to facilitate virulence relationships with eukaryotes.

Background

Over a dozen distinct families of transport systems that catalyse translocation of proteins across cytoplasmic membranes have been recognized, characterized and classified (Saier, 2000). Of these, most are restricted to bacteria. Only a few are either eukaryote-specific or ubiquitous. Among the most complex of the bacterial systems are the type IV protein secretion pathway systems of Gram-negative bacteria [members of the IVSP family; transporter classification (TC) number 3.A.7] which have long been recognized as the systems responsible for the exchange of genetic material in the process known as bacterial conjugation (Bohne et al., 1998; Dreiseikelmann, 1994; Frost et al., 1994). More recently, great interest in type IV secretion has resulted from the discovery that agrobacteria use a related system to transfer a segment of their nucleic acid, called T-DNA, together with specific proteins, to plant host cells in a process that results in cancerous transformation (Christie, 1997; Christie & Vogel, 2000, 2001; Gelvin, 2000; Kado, 1994; Lessl et al., 1992; Shirasu et al., 1994; Tzfira et al., 2000). Ten proteins have been directly implicated in the DNA transfer process (Berger & Christie, 1994; Burns, 1999; Christie, 1997; de la Cruz & Lanka, 1998; Kuldau et al., 1990; Thorstenson et al., 1993; Winans et al., 1996). A related export system mediates pertussis toxin secretion (Covacci & Rappuoli, 1993; Nicosia et al., 1986; Weiss et al., 1993) and the H. pylori and Legionella pneumophila genomes encode genes related to IVSP system constituents that are important for virulence (Covacci et al., 1997, 1999; Ramarao et al., 2000; Segal et al., 1998; Tummuru et al., 1995; Vogel et al., 1998). The agrobacterial system...
has recently been shown to translocate free proteins as well as nucleoprotein complexes (Vergunst et al., 2000).

In addition to VirB2–11 homologues, VirD4 homologues are thought to play a role as coupling proteins, linking DNA–protein substrates to the membrane pore (Hamilton et al., 2000). The X-ray structure of the VirD4 homologue, TrwB from plasmid R388, has recently been solved (Gomis-Ruth et al., 2001). Homologues of this protein are found in H. pylori and R. prowazekii, organisms that do not possess a full complement of the VirB homologues (C. Baron, personal communication). Although VirD4 and several other peripheral constituents are important for the secretory process, we will nevertheless focus on the VirB2–VirB11 constituents that are believed to play a direct role in translocation by providing the structural constituents of the secretory apparatus.

Although previous efforts have resulted in the identification of common type IV secretory system constituents in a variety of Gram-negative bacteria (Christie, 2001; Krause et al., 2000; Li et al., 1999), no previously published work has described systematic phylogenetic analyses of these proteins. We have therefore initiated a comprehensive analysis of these proteins using the agrobacterial VirB system as a starting point. We identify all recognizable homologues of the VirB2–11 proteins, characterize these homologues with respect to organismal source and size, and align the sequences for the purposes of (1) constructing phylogenetic trees, (2) deriving mean hydropathy and amphipathicity plots that lead to estimates of topology, (3) determining regions of high conservation, and (4) identifying well-conserved residues that are likely to be of structural or functional significance.

Only representative summary tables and data figures are presented to document our most important conclusions. The complete list of protein homologues in the ten families (the VirB2 to VirB11 families), including protein abbreviations, database descriptions, organismal sources, protein sizes and database accession numbers that allow easy access to the sequences, can be found as supplementary data (see http://mic.sgmjournals.org). Both the clustalx and tree programs were used to derive phylogenetic trees although only clustalx-derived trees are presented. Both alignments were also used to derive mean hydropathy, similarity and amphipathicity plots as outlined previously (Kyte & Doolittle, 1982; Le et al., 1999). A sliding window of 21 residues was generally used, although for identification of amphipathic β structure a sliding window of 7 or 9 residues was sometimes used with the tree program (Le et al., 1999) or the averag program (Zhai & Saier, 2001a). For calculation of mean amphipathicity, the angle used was 100° for z structure or 180° for β structure (Le et al., 1999; Zhai & Saier, 2001a). Topology was estimated using the tmapred (Hofmann & Stoffel, 1993) and das (Cserzo et al., 1997) programs on individual proteins (Zhai & Saier, 2001b) as well as by viewing the mean hydropathy plots.

In the case of the large VirB11 family of (putative) ATPases, the phylogenetic trees were derived both with the complete sequences and with the most conserved regions of these proteins. These latter regions were selected by visual inspection and were spliced according to the alignment of their full sequences generated with clustalx. This method obviates the problem of artifactualy distorted branch lengths due to association with non-homologous or poorly aligned sequences.

The organisal abbreviations for organisms that clearly encode parts of IVSP systems include (1) Atu, A. tumefaciens; (2) Bab, Brucella abortus; (3) Bhe, Bartonella henselae; (4) Bpe, Bordetella pertussis; (5) Bs, Brucella suis; (6) Eae, Enterobacter aerogenes; (7) Eco, Escherichia coli; (8) Hpy, H. pylori; (9) Lpn, L. pneumophila; (10) Reu, Ralstonia eutropha; (11) Ret, Rhizobium etli; (12) Rsp, Rhizobium sp.; (13) Rpr, Rickettsia prowazekii; (14) Sii, Salmonella typhi; (15) Sry, Salmonella typhimurium; (16) Xfa, Xylella fastidiosa. Plasmids encoding such systems include RK2, RP4 and IncF.

**Computer methods**

The VirB2–VirB11 proteins used as query sequences for identification of protein homologues were from Agrobacterium tumefaciens AGRT9 (SWISS-PROT) (plasmid pTi15955). The psi-blast program (Altschul et al., 1997) was used with iterations to convergence to identify all homologues exhibiting sufficient sequence/motif similarity to be identified by this method. All proteins retrieved were checked for consistency with respect to available functional/structural information. Additionally, sequence-divergent proteins previously identified and reported to be homologous (or functionally analogous) by Christie (1997) were screened for homologues using the same procedure. Recursive blast (Jack et al., 2000) was also used to identify distant homologues. The protein homologues within these ten protein families, retrieved by these methods, are shown as supplementary data.

Two programs were used to align the protein sequences retrieved as described above: clustalx (Thompson et al., 1997) and tree (Feng & Doolittle, 1990) (see Young et al., 1999 for comparison and evaluation of these and other available programs). The gap penalty and gap extension values used with the clustalx program were 10 and 0·1, although other combinations were tried. This set of values consistently gave similar phylogenetic trees to those calculated with values of 8 and 2, respectively. Default parameters were used for the tree program. The multiple alignments from which the results are derived, as well as many of the derived figures are not presented here, can be found at http://mic.sgmjournals.org. Both the clustalx and tree programs were used to derive phylogenetic trees although only clustalx-derived trees are presented. Both alignments were also used to derive mean hydropathy, similarity and amphipathicity plots as outlined previously (Kyte & Doolittle, 1982; Le et al., 1999). A sliding window of 21 residues was generally used, although for identification of amphipathic β structure a sliding window of 7 or 9 residues was sometimes used with the tree program (Le et al., 1999) or the averag program (Zhai & Saier, 2001a). For calculation of mean amphipathicity, the angle used was 100° for z structure or 180° for β structure (Le et al., 1999; Zhai & Saier, 2001a). Topology was estimated using the tmapred (Hofmann & Stoffel, 1993) and das (Cserzo et al., 1997) programs on individual proteins (Zhai & Saier, 2001b) as well as by viewing the mean hydropathy plots.
VirB protein homologues: overview

Tables 1, 2 and 3 summarize the data currently available concerning homologues of the VirB2–11 proteins of *A. tumefaciens*. Between 11 and 89 homologues were identified for each of the 10 recognized protein constituents of the IVSP family. Thus, 11 VirB7 homologues, 14 VirB5 and VirB6 homologues, 17 VirB3 homologues, 18 VirB2 homologues, 19 VirB8 and VirB10 homologues, 22 VirB9 homologues, 29 VirB4 homologues and 89 VirB11 homologues were identified (Table 1). VirB11 ATPases function in many capacities other than type IV secretion (see below) explaining the large number of these homologues. Because VirB7 is the smallest protein of the IVSP, identification of its homologues proved difficult, possibly accounting for the small number of these homologues identified. VirB4 is the largest constituent, partially accounting for the large number of homologues found. However, like VirB11 ATPases, VirB4 ATPase homologues function in capacities other than IVSPs. All the remaining constituents have about the same numbers of homologues in the databases (14–22) and all of these proteins are of an intermediate size (100–400 residues).

Four methods were used to estimate the topologies of representative proteins of each of the IVSP. First, Tmpred was used on several representative proteins. Second, the DAS program was used, analysing the same proteins. Third, the numbers of putative TMSs (transmembrane segments) reported by Dang & Christie (1997) or Das & Xie (1998; 2000) are reported. Finally, topology was estimated from the mean hydropathy plots derived in this study. At least one hydrophobic region, sometimes possibly corresponding to an N-terminal, cleavable signal sequence, was identified in each family of IVSP constituents. For example, VirB2 homologues have three putative TMSs, VirB4 homologues exhibit two to four putative TMSs, VirB6 homologues may have five or six TMSs, VirB9 homologues have one or two putative TMSs and all other IVSP homologues probably exhibit one putative TMS (Table 1). It should be noted that these predictions are dependent on the methods used (see Table 1).

Functions suggested for these constituents are shown in Table 2. VirB4 and VirB11 are ATPases. VirB11 may function in biogenesis/assembly of the transporter complex while VirB4 probably energizes the translocation process. VirB6 and VirB4 may together comprise the inner membrane channel. Alternatively, VirB8, B9 and B10, which we suggest form a transperiplasmic channel as well as the outer membrane pore, could comprise the cytoplasmic channel. VirB7 is a lipoprotein that may anchor and stabilize the VirB8–10 complex and/or the pilus to the outer membrane (Sagulenko *et al.*, 2001). VirB2, B3 and B5 may be constituents of a pilus or pilus-like structure (Christie, 2001; Sagulenko *et al.*, 2001; Shirasu & Kado, 1993).

Our phylogenetic analyses revealed 10 overall clusters of sequence-related proteins (clusters 1–10; see Tables 1 and 3). All 10 IVSP constituents were found in clusters 1, 3, 4, 5, 7 and 8, suggesting that all constituents of at least one represented system in each of these clusters have been identified and sequenced. However, nine constituents of cluster 6, five constituents of clusters 2 and 9, and only one of cluster 10 are represented. Proteins in these clusters may serve functions not requiring all protein components of an IVSP.

### Table 1. VirB proteins and their homologues: proposed topologies, sizes and organismal distributions

<table>
<thead>
<tr>
<th>Vir protein</th>
<th>No. proteins*</th>
<th>Mean size† (aa)</th>
<th>Size range (aa)</th>
<th>No. TMS§</th>
<th>Cluster represented§</th>
<th>No. clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>B2</td>
<td>18</td>
<td>119±16</td>
<td>96–145</td>
<td>(3)(2)(3)</td>
<td>+ + + + + + + + +</td>
<td>8</td>
</tr>
<tr>
<td>B3</td>
<td>17</td>
<td>103±7</td>
<td>92–116</td>
<td>(2 or 3)(1)(1)(1)</td>
<td>+ + + + + + + + +</td>
<td>8</td>
</tr>
<tr>
<td>B4</td>
<td>29</td>
<td>815±67</td>
<td>662–893</td>
<td>(4)(2)(4)(3)</td>
<td>+ + + + + + + + +</td>
<td>10</td>
</tr>
<tr>
<td>B5</td>
<td>14</td>
<td>226±29</td>
<td>188–269</td>
<td>(1–3)(2)(1)(1)</td>
<td>+ + + + + + + + +</td>
<td>7</td>
</tr>
<tr>
<td>B6</td>
<td>14</td>
<td>369±74</td>
<td>295–372</td>
<td>(7)(5–6)(5)(3–6)</td>
<td>+ + + + + + + +</td>
<td>7</td>
</tr>
<tr>
<td>B7</td>
<td>11</td>
<td>38±9</td>
<td>47–69</td>
<td>(1)(1)(1)(1)</td>
<td>+ + + + + + + + +</td>
<td>7</td>
</tr>
<tr>
<td>B8</td>
<td>19</td>
<td>246±33</td>
<td>202–260</td>
<td>(1)(2)(1)(1)</td>
<td>+ + + + + + + + +</td>
<td>8</td>
</tr>
<tr>
<td>B9</td>
<td>22</td>
<td>306±81</td>
<td>328–522</td>
<td>(2)(2)(1)(0 or 2)</td>
<td>+ + + + + + + + +</td>
<td>8</td>
</tr>
<tr>
<td>B10</td>
<td>19</td>
<td>411±42</td>
<td>336–483</td>
<td>(3)(1–3)(1)(1)</td>
<td>+ + + + + + + + +</td>
<td>9</td>
</tr>
<tr>
<td>B11</td>
<td>89</td>
<td>436±102</td>
<td>252–721</td>
<td>(1)(2)(0)(1)</td>
<td>+ + + + + + + + +</td>
<td>8</td>
</tr>
</tbody>
</table>

* No. protein homologues identified.
† Mean protein size in number of amino acyl residues ± standard deviation.
‡ No. TMSs predicted by using (1) Tmpred, (2) DAS, (3) method reported by Dang & Christie (1997) or Das & Xie (1998; 2000), and (4) as determined by mean hydropathy for representative sequence divergent homologues, respectively.
§ See phylogenetic trees and Table 3 for identification of individual proteins within each cluster.

3203
### Table 2. Predictions of functions and subcellular locations for constituents of the VirB IVSP complex

Cyto, cytoplasm; IM, inner membrane; P, periplasm; OM, outer membrane.

<table>
<thead>
<tr>
<th>Function*</th>
<th>Location†</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>VirB2</td>
<td>Major pilus subunit</td>
<td>Exported; envelope-associated</td>
</tr>
<tr>
<td>VirB3</td>
<td>Minor pilus subunit?</td>
<td>Exported; envelope-associated</td>
</tr>
<tr>
<td>VirB4</td>
<td>ATPase (for nucleoprotein transporter?) (IM pore?)</td>
<td>Cyto-IM-P</td>
</tr>
<tr>
<td>VirB5</td>
<td>Minor pilus subunit; pilus tip or pilus base?</td>
<td>Exported; envelope-associated</td>
</tr>
<tr>
<td>VirB6</td>
<td>IM pore or T-pilus regulator</td>
<td>IM (associated with VirB4?)</td>
</tr>
<tr>
<td>VirB7</td>
<td>Lipoprotein; interacts with VirB9 and the pilus</td>
<td>OM; periplasmic face</td>
</tr>
<tr>
<td>VirB8</td>
<td>IM-P-OM pore complex</td>
<td>IM-P-OM</td>
</tr>
<tr>
<td>VirB9</td>
<td>IM-P-OM pore complex</td>
<td>IM-P-OM</td>
</tr>
<tr>
<td>VirB10</td>
<td>IM-P-OM pore complex</td>
<td>IM-P-OM</td>
</tr>
<tr>
<td>VirB11</td>
<td>ATPase (for IVSP complex assembly?)</td>
<td>Cyto</td>
</tr>
</tbody>
</table>

*Functional predictions are based on the published literature as well as sequence analyses.
†Proposed subcellular locations are based on the published literature and our secondary structure and topological analyses.

### Table 3. Homologues of IVSP constituents in the ten phylogenetic clusters identified in this study

<table>
<thead>
<tr>
<th>Cluster no. (protein group or plasmid)…</th>
<th>7, 8 (VirB)</th>
<th>1 (IncP)</th>
<th>3 (LvhB)</th>
<th>4 (IncN)</th>
<th>5 (IncW)</th>
<th>6 (Ptl)</th>
<th>9 (IncF)</th>
<th>2 (Hpy)</th>
<th>10 (Gram + bacteria)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VirB3</td>
<td>TrbD</td>
<td>LvhB3</td>
<td>TrbA</td>
<td>TrwM</td>
<td>PtlB</td>
<td>TraL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>VirB4</td>
<td>TrbE</td>
<td>LvhB4</td>
<td>TrbB</td>
<td>TrwK</td>
<td>PtlC</td>
<td>TraC</td>
<td>VirB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VirB5</td>
<td>TrbJ</td>
<td>LvhB5</td>
<td>TrbC</td>
<td>TrwJ</td>
<td>PtlD</td>
<td>TraE</td>
<td>Cag, PicB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VirB6</td>
<td>TrbL</td>
<td>LvhB6</td>
<td>TrbD</td>
<td>TrwI</td>
<td>PtlI</td>
<td></td>
<td>ComB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VirB7</td>
<td>TrbK</td>
<td>LvhB7</td>
<td>TrbN</td>
<td>TrwH</td>
<td>PtlH</td>
<td></td>
<td>ComB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VirB8</td>
<td>TrbF</td>
<td>LvhB8</td>
<td>TrbE</td>
<td>TrwG</td>
<td>PtlE</td>
<td></td>
<td>Orf, ComB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VirB9</td>
<td>TrbG</td>
<td>LvhB9</td>
<td>TrbO</td>
<td>TrwF</td>
<td>PtlF</td>
<td></td>
<td>Cag, ComB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VirB10</td>
<td>TrbI</td>
<td>LvhB10</td>
<td>TrbF</td>
<td>TrwE</td>
<td>PtlG</td>
<td>TraB</td>
<td>ComB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VirB11</td>
<td>TrbB</td>
<td>LvhB11</td>
<td>TrbG</td>
<td>TrwD</td>
<td>PtlH</td>
<td></td>
<td>VirB, Orf, IIIISP</td>
<td></td>
</tr>
</tbody>
</table>

Table 3 presents the names of representative proteins in each of the 10 clusters. Clusters 7 and 8 include VirB proteins of *Agrobacterium* and of *Bartonella*, respectively. Clusters 1, 4 and 5 include the Trb proteins of plasmid IncP, the Tra proteins of plasmid IncN and the Trw proteins of IncW, respectively. Cluster 6 includes the Ptl pertussis toxin export proteins of *Bor. pertussis*. Only a VirB5 homologue is missing in this system. Because the Ptl system does not transport DNA and functions exclusively to export a toxin, this deficiency may have functional significance. All other clusters (clusters 2, 9 and 10) are not represented in several of the phylogenetic trees. Proteins included in these clusters may be constituents of incompletely sequenced IVSP systems or constituents of other types of systems. The latter possibility, clearly valid for some constituents...
such as the VirB4 and VirB11 ATPase homologues, is consistent with the notion that IVSP systems were originally constructed in part from pre-existing proteins that were constituents of other types of systems (see Conclusions).

A list of all of the proteins identified in each of the ten protein families is available at http://mic.sgmjournals.org. A few of these entries either are putative fragments or are homologues of strikingly dissimilar size as compared with the other homologues. These few proteins were excluded from our analyses. Table 1 should be used as a guide for the analyses reported below dealing with the ten recognized constituents of the IVSP family.

VirB2
Eighteen VirB2 homologues of 96–154 aa were identified and aligned. The mean hydropathy, mean amphipathicity and mean similarity plots, and phylogenetic trees for members of the VirB2 family as well as for the VirB3–11 families can be viewed at http://mic.sgmjournals.org. Three peaks of mean hydropathy are observed for the VirB2 proteins. The first of these exhibits lower hydropathy and lower sequence similarity than the second two. The functionally most important parts of these proteins are therefore predicted to include putative TMSs 2 and 3.

Mean amphipathicity plots revealed that when the angle was set at 100°, as is appropriate for an α-helix, peak 2 is preceded by a region of striking amphipathicity, while a region of lesser amphipathicity separates putative TMSs 2 and 3. With the angle set at 180°, as for a β-strand, the region immediately preceding putative TMS1 exhibits a strong peak. This information suggests that VirB2 homologues exhibit β structure in their N-terminal regions but α structure in remaining parts.

No fully conserved residues were identified in the VirB2 multiple alignment. However, regions of striking sequence similarity were observed in putative TMSs 2 and 3. At several positions within these two secondary structural elements, only hydrophobic residues were found. Well-conserved glycyl residues were also present in both regions, suggesting structural flexibility. At the end of helix 3, a well-conserved hydroxy amino acid (serine or threonine) was found. The results suggest that although no residue is essential for a single function shared by all of these proteins, certain residue types are important to maintain a requisite structure/function relationship.

The phylogenetic tree for the VirB2 family shows that eight major clusters are present; clusters 2 and 10 are missing (see Tables 1 and 3). The clustering of proteins within each major cluster is undoubtedly of phylogenetic significance. This is particularly relevant for large clusters 1, 6 and 9. Other clusters or branches are either represented by a single protein (clusters 3, 4, 5 and 8) or by a pair of close orthologues (cluster 7).

VirB3
Seventeen VirB3 homologues were identified. These short proteins exhibit an extended N-terminal hydrophobic region of about 40 residues which could possibly form one or two TMSs. This hydrophobic region is preceded by the region of greatest sequence similarity. Striking regions of amphipathicity both precede and follow the hydrophobic region when the angle was set at 100°. No region of well-conserved amphipathicity was observed when the angle was set at 180°. The multiple alignment for the VirB3 family proteins revealed a well-conserved (GAT)L(ST)RP motif (alternative residues at a single alignment position are in parentheses) just preceding the region of strong hydrophobicity. Additionally, a glycyl residue near the N-terminal region of hydrophobicity was fully conserved in all proteins except TrbD from plasmid pM3 where a prolyl residue was found. All other well-conserved regions were strongly hydrophobic, but no one residue predominated. These observations suggest a primary structural role for this region.

The phylogenetic tree for VirB3 homologues revealed that the proteins in clusters 1, 6 and 9 exhibit relative distances from each other similar to those observed for the corresponding proteins in the VirB2 tree. Also, the Esc. coli Tra and Trw proteins of clusters 4 and 5 (Table 3) are more closely related to each other than to other homologues. Thus, it appears that VirB2 and B3 homologues have evolved in parallel without shuffling of constituents between systems.

VirB4
VirB4 homologues are large dimeric proteins of about 800 residues per subunit (Dang et al., 1999). They exhibit ATPase activity and may energize transport (Shirasu et al., 1994). Each subunit exhibits two periplasmic domains, traversing the cytoplasmic membrane multiple (probably four) times (Table 1; Dang & Christie, 1997). Twenty-nine homologues were identified. The hydrophobicity and similarity plots reveal that the N- and C-terminal thirds of these proteins exhibit properties that are typical of hydrophilic proteins. The four putative TMSs are localized to a region near the centres of these proteins. These homologues exhibit nearly uniform sequence similarity throughout. With the angle set at 100°, several regions exhibit striking amphipathicity. The largest is between residue positions 180 and 220 in the multiple alignment, but additional peaks are observed at alignment positions 390 and 800. With the angle set at 180°, a single large peak of mean amphipathicity is observed at position 360.

The multiple alignment of the VirB4 homologues revealed several regions of sequence conservation separated by gaps. Most striking was the well-conserved BBGX(ST)G(AS)GK(ST)XBB motif at alignment positions 710–725 where B is a hydrophobic residue and X is any residue. This motif presumably serves for nucleotide binding. Adjacent to this motif is another
well-conserved region: BBBDXDXGX<sub>2</sub>BNPLB (alignment positions 764–802). Other regions (positions 980–1050) also proved to be highly conserved. Thus, these strongly conserved residues occur in the C-terminal portions of the VirB4 homologues where the enzyme active sites responsible for substrate recognition are likely to be found.

The phylogenetic tree for VirB4 homologues is most revealing. First, clusters 1, 6 and 9 are strikingly similar to those shown for VirB2 and VirB3 homologues. Second, the Tra and Trw proteins (clusters 4 and 5) are again adjacent to each other. Third, unexpectedly, these two proteins cluster loosely with cluster 6, a fact that may reflect the higher phylogenetic resolution possible with proteins of long sequence. Finally, clusters 3, 4, 5, 7 and 8 each generally consists of a single protein (or a pair of orthologues) as observed for the VirB2 and VirB3 family trees.

The only significant difference between this tree and those for the VirB2 and VirB3 families is the presence of clusters 2 and 10, which were lacking from the latter trees. Cluster 2 includes proteins exclusively from <i>H. pylori</i> while cluster 10 includes proteins exclusively from low G+C Gram-positive bacteria. Since cluster 10 is found only in the VirB4 family, it can be concluded that the homologous Gram-positive bacterial ATPases serve a function that is very different from those exhibited by their homologues in Gram-negative bacteria. At least some of the <i>H. pylori</i> homologues, of widely differing sequences, are likely to serve non-IVSP functions that, however, may include other VirB homologues.

VirB6

Like the VirB5 family, the VirB6 family consists of 14 sequenced proteins. These proteins may associate with their VirB4 homologues and form the transcytoplasmic membrane pore. Alternatively, they may stabilize the T-pilus (Hapfelmeier et al., 2000). These proteins are of about 300 residues but exhibit considerable size variation (Table 1).

Mean hydropathy and similarity plots for the VirB6 homologues reveal two well-conserved N-terminal peaks of hydrophobicity, followed by a poorly conserved hydrophilic region, a second hydrophobic region of about 40 residues (probably two close TMSs separated by a β turn) and two final hydrophobic peaks, each about equidistant from the other and from the preceding region of hydrophobicity. They are somewhat less well conserved than the previous ones. There are thus six putative TMSs (Table 1).

VirB6 homologues are relatively poorly conserved with no fully conserved residues. Most of the well-conserved residues either are hydrophobic in nature or are glycyl residues within the transmembrane regions. However, the mean amphipathicity plots revealed that preceding, in between, and following all putative TMSs (except between TMSs 3 and 4, and between TMSs 5 and 6) are regions of very strong amphipathicity when the angle is set at 100° for α helix. We therefore predict that VirB6 proteins are largely α-helical, both in the membrane and in the inter-TMS loop regions.

The phylogenetic tree for the VirB6 family resembles those discussed previously, although PtID Bpe (cluster 6) is more distant from the other members of its cluster than expected. The Tra and Trw proteins of <i>Esc. coli</i> again cluster loosely together. Clusters 2, 9 and 10 are not represented (Table 2). This is surprising in view of the large number of cluster 9 homologues. This fact leads us to suggest that cluster 9 VirB homologues function by mechanisms different from those used by most IVSP family members. Alternatively, some of the IncF-like plasmid transfer genes may have not been sequenced, or some of the constituents may exhibit too great a degree of sequence divergence to be recognized.

VirB7

The <i>A. tumefaciens</i> outer membrane VirB7 exo-lipo-protein is reported to be associated with both the VirB9 protein and the T-pilus (Sagulenko et al., 2001). The VirB7 family consists of 11 proteins, all of about 55 residues. An additional potential VirB7 homologue was reported by Ramarao et al. (2000), but it differs in size from other VirB7 homologues and shows insufficient sequence similarity to establish homology. Therefore, this protein was omitted from our studies.

The mean hydrophobicity plot of the 11 established members of the VirB7 family reveals a single N-terminal TMS of less than 20 residues followed by a 30 residue hydrophilic peptide. This hydrophilic region did not
exhibit a striking amphipathic character when the angle was set to 100°, but its C-terminal portion showed three sharp peaks when the angle was set at 180° and the sliding window was set at seven residues. This suggests that the C-terminal regions of VirB7 lipoproteins may contain substantial β structure.

Multiple alignments revealed positively charged residues preceding and a fully conserved cysteyl residue following the N-terminal leader sequence motif as expected for these lipoproteins. A well-conserved motif including this cysteyl residue is (LIV)(SAG)(GA)C. The hydrophilic C-termini of these proteins are proline rich but otherwise poorly conserved. The phylogenetic tree resembles those described previously for other VirB family homologues in virtually all respects, a fact that is surprising in view of the small sizes of these homologues.

**VirB8**

The VirB8 family consists of proteins of approximately 250 residues with a cytoplasmic, hydrophilic N-terminus, a single TMS and a large extracellular C-terminal domain. The strongly hydrophilic extracellular C-terminal region, of fairly uniform sequence similarity, lacks appreciable amphipathicity when analysed with an angle of 100°. However, the last 100 residues exhibit several sharp peaks of amphipathicity when analysed with an angle of 180°, suggesting a β sheet region that may be associated with, or inserted into, the outer membrane.

Except for well-conserved R/K residues preceding the putative TMS and one or two well-conserved prolyl residues following this putative TMS, the N-terminal region shows little sequence conservation. However, in the C-terminal portion we found a Y(VI)X,#RE motif that is fully conserved except in the Trb proteins (cluster 1). Further downstream of this conserved motif are well-conserved hydroxy amino acids (ST), aromatic amino acids (YF) and hydrophobic residues (LIV) interspersed with hydrophilic residues. These latter regions are responsible for the β-structure amphipathicity. The phylogenetic tree shown in Fig. 1 reveals clustering of VirB8 family proteins in accordance with expectation assuming that IVSP systems have evolved without shuffling of constituents between systems.

**VirB9**

The VirB9 family includes 22 members of about 300 residues with one large and one small peak of hydrophobicity, possibly corresponding to one or two TMSs. These two peaks of hydrophobicity exhibit striking sequence similarity. They are followed by a poorly conserved but strikingly hydrophilic region of about 150 residues followed by a well-conserved, more hydrophobic C-terminal sequence of about 70 residues. The amphipathicity plots reveal short stretches of amphipathicity when the angle is set at 180° in this last mentioned region. Moreover, the multiple alignment reveals several short stretches of six to nine residues with alternating hydrophilic and hydrophobic residues, typical of β-barrel porin proteins. These are often separated by regions of poorer sequence conservation. These observations are consistent with an association of the C-termini of these proteins with the outer membrane in a β-pleated sheet structure, possibly comprising all or part of an outer membrane pore structure.
Fig. 2 shows the VirB9 family tree. There are no surprises with respect to the clustering of most of the proteins. However, four *H. pylori* proteins comprise cluster 2. Cluster 2 is represented in only five of the VirB families. The two VirB9 homologues from *Ric. prowazekii* cluster loosely with the LvhB protein of *L. pneumophila*. It is of considerable interest that *Ric. prowazekii* homologues of VirB4, B8, B9, B10 and B11 were identified, but homologues of the other VirB proteins were not found in this organism. A similar scenario is observed for *H. pylori* but not for *L. pneumophila*. Since the genomes of *H. pylori* and *Ric. prowazekii* have been completely sequenced, we suggest that these proteins may serve common functions in virulence that do not require a complete IVSP system and possibly comprise part of a different type of transporter. A unified function for the latter four of these five VirB constituents can be suggested. It is interesting to note that *Ric. prowazekii* displays only one VirB4, VirB8, VirB10 or VirB11 homologue, but two sequence-divergent VirB9 homologues. Because their sequences are so different, it seems unlikely that the latter two paralogues serve the same function.

**VirB10**

Nineteen VirB10 family members were identified. A single hydrophilic sequence is followed by a strongly hydrophilic stretch of nearly 200 residues, and then comes a more hydrophilic region of about 150 residues that is well conserved. Amphipathicity plots revealed that this last C-terminal segment exhibits short sequences of probable amphipathic $\beta$ structure. These, like the C-terminal regions of the VirB8 and VirB9 proteins may be inserted into the outer membrane. These three proteins have such similar structures and sizes that it would be reasonable to suggest that they function together and form a heterooligomeric structure. The N-terminal hydrophobic regions are suggested to span the cytoplasmic membrane; the central portions span the periplasm, and the C termini are associated with or embedded in the outer membrane. They could provide a continuous transenvelope channel (see Conclusions).

The multiple alignment of the VirB10 homologues supports this supposition. Thus, their N termini, preceding the single TMS, are cationic. This is followed by a region of poor conservation (the proposed transperiplasmic region), and finally, in the C-termini one finds alternating hydrophilic and hydrophobic residues. Within this C-terminal putative $\beta$-sheet region are four fully conserved and an additional nine well-conserved glycy1 residues. They occur at irregular intervals of between 3 and 17 residues, most at odd-numbered intervals. Well-conserved cationic residues (R and K) are also found. Finally, five well-conserved blocks of three to five hydrophilic residues as well as occasional aromatic residues and proline can be found in this C-terminal domain. These observations undoubtedly have both structural and functional significance in the formation of a transenvelope structure.

The phylogenetic tree for the VirB10 proteins (Fig. 3) shows a configuration that illustrates the proposed co-evolution of the 10 VirB family homologues. However, the loose clustering of ComB3 Hpy with VirB10 Rpr suggests that these two proteins may serve a similar function. ComB3 Hpy is assigned to cluster 2, corresponding to the clusters 2 found in previously mentioned phylogenetic trees.
VirB11

The VirB11 family is by far the largest of the ten VirB families. We identified 89 sequenced homologues, all derived from prokaryotes. However, in contrast to all but one of the other VirB families, the VirB11 family includes many proteins from Gram-positive bacteria, and in contrast to all other VirB families, the VirB11 family includes proteins from archaea. Also represented are large numbers of VirB11 homologues from a diversity of bacteria not represented in the other phylogenetic trees. Many of the bacterial proteins are from photosynthetic organisms. VirB11-type ATPases function in capacities other than as essential components of IVSPs. Thus, (1) some are constituents of the main terminal branch of the type II general secretory pathway; (2) others function in fimbrium (pilus) assembly; (3) still others function as constituents of natural-competence-related DNA-uptake systems; and finally (4) others are believed to play a role in bacterial twitching motility.

The mean hydropathy and similarity plots for the VirB11 proteins reveal a strongly hydrophilic character with the hydrophilic domains showing the greatest sequence similarity. The mean hydropathy plot indicates the presence of a single region sufficiently long and hydrophobic to traverse the membrane, but this could be the hydrophobic core of a water-soluble protein.

Mean amphipathicity plots reveal three broad regions of putative amphipathic $\beta$ strand (alignment positions 80–130, 200–270 and 320–390). It seems likely that VirB11 homologues include considerable $\beta$ structure.

The multiple alignment reveals several well-conserved motifs. At position 516–528 is a well-conserved motif: $(B)_{3}(AST)G^*(PG)T(AGS)SG^*K^*(ST)T$ (asterisks indicate fully conserved residues). A second well-conserved motif, centred at alignment position 550, is $B_{3}(ST)BED(ST)EB$. The ED motif is almost fully conserved. A third well-conserved motif is centred at alignment position 598. This large region exhibits the motif $B(ST)GX_{3}B_{3}(KR)B_{3}RX_{3}P(DE)XB_{3}GE^*BR$. Finally, a well-conserved TGH motif occurs at alignment positions 624–626.

Phylogenetic trees were derived both for the full length VirB11 proteins and for the best conserved regions of these proteins (about 120 residues; alignment positions 512–628 in the full-length sequence alignment). Branching patterns in the two trees proved to be in excellent agreement. The former tree is reproduced in Fig. 4. All of the proteins that comprise the IVSP family constituents are found in cluster 8. Moreover, within cluster 8, eight of the ten subclusters identified in the other VirB family trees are found with configurations as expected assuming that these VirB11 homologues co-evolved in parallel with the other VirB proteins. This observation provides convincing evidence for the lack of appreciable shuffling of protein constituents between IVSP systems during their evolution. All remaining clusters in the VirB11 tree (clusters 1–7, 9 and 10) are not represented in the other VirB family trees. The proteins comprising most or all of these clusters may serve a unified function in energizing export and/or promoting assembly of extracellular protein complexes. It is suggested that VirB11 was ‘borrowed’ from pre-existing systems for the assembly of more recently arising IVSP systems.

Conclusions

We have analysed 10 families of IVSP constituents with respect to their sequences, structures and phylogenetic relationships. The results can be interpreted in terms of...
a transenvelope structure as depicted in Fig. 5 and summarized as follows.

(1) The VirB proteins vary tremendously in size, from very small (VirB7 is about 55 residues), moderately small (VirB3 and B2 are about 100 and 120 residues, respectively), moderate in size (VirB5, B6, B8, B9, B10 and B11 are 220–400 residues), to large (VirB4 is about 800 residues) (Table 1). These size differences impact on the reliability of phylogenetic data derived from analyses of their sequences.

(2) These proteins may be primarily cytoplasmically localized (VirB4 and VirB11), and/or they may be localized to the inner membrane as integral membrane proteins (VirB4 and VirB6). Three of them (VirB8, B9 and B10) are predicted to traverse both membranes and the periplasm with their N termini in the cytoplasm and their C termini embedded in the outer membrane (Kumar & Das, 2001). VirB7 is an outer membrane lipoprotein that may anchor both the VirB8–B9–B10 channel and the pilus to the outer membrane (Sagulenko et al., 2001). VirB2, B3 and B5 may be constituents of the pilus (or pilus-like organelle) that mediates cell–cell interactions or functions as a piston to drive protein export (Christie, 2001).

(3) The two putative cytoplasmic constituents are ATPases; one (VirB11) may function in biogenesis of the IVSP system (Yeo et al., 2000) while the other (VirB4) may energize transport of the substrate protein or nucleoprotein complex (Christie, 2001). The six-TMS integral inner membrane VirB6 protein may, together with VirB4, comprise an oligomeric pore, providing the protein translocation pathway across the inner mem-
brane, but it may also play a role in pilus stabilization. The outer-membrane-associated proteins VirB8, B9 and B10 may comprise the envelope pore, and VirB7, a small outer-membrane lipoprotein, may stabilize the VirB8–B9–B10 complex and/or anchor it and the pilus to a specific region of the outer membrane (Sagulenko et al., 2001). VirB8, B9 and B10 appear to each exhibit a single N-terminal inner transmembrane helix, a central transperiplasmic α-helical region and a C-terminal β-structured domain that may interact with and/or insert into the outer membrane of the Gram-negative bacterial envelope. Thus, these three proteins may comprise or interact with the inner-membrane channel formed by VirB4–VirB6 to allow construction of a continuous channel through the entire envelope (the inner membrane, the periplasm and the outer membrane). Such a structure could be either static or dynamic. In the latter case, its formation could be induced by the presence of a substrate protein as has been demonstrated for ABC (type I secretory pathway) protein export (Binet et al., 1997). In either case, however, the conduit could allow transport of macromolecules directly across the entire cell envelope in a single energy-coupled step.

(4) All 10 constituents of the IVSP family transporters are found in at least five sequence-divergent systems. These systems, in addition to the agrobacterial and bartonellar VirB systems (phylogenetic clusters 7 and 8, respectively), include the Trb proteins of plasmid IncP (cluster 1), the LvhB proteins of L. pneumophila (cluster 3), the Tra proteins of plasmid IncN (cluster 4) and the Trw proteins of plasmid IncW (cluster 5). None of the ten VirB proteins are found in the Pti system of Bor. pertussis (Table 2), a system involved in export of Bordetella protein toxin(s). Only VirB5 was not identified in this system, and VirB5 homologues have been proposed to be adhesin-like proteins localized to the pilus (fimbrial) tip. Alternatively, they may associate with the pilus in another capacity (see Table 1). Other systems such as the Tra system of plasmid IncF (cluster 9) seem to lack several constituents, possibly because of incomplete plasmid sequencing. The VirB8, B9 and B10 proteins, and possibly VirB11 as well, that comprise cluster 2 probably form a structure that functions in a unified capacity, possibly in virulence and/or competence in H. pylori (Ramarao et al., 2000) and Ric. prowazekii by a mechanism that does not utilize a complete IVSP system. This suggestion was recently reinforced by a report demonstrating the presence of an operon in Wolbachia species that encodes VirB8, 9, 10 and 11 as well as VirB4 but not other IVSP constituents. This operon is required for arthropod sexual alteration as has recently been suggested (Kumar & Das, 2001), they could together serve a variety of functions in addition to type IV secretion. Finally, the VirB4 ATPase is the only constituent found in cluster 10. These proteins are derived from Gram-positive bacteria where the ATPase must energize a process unrelated to type IV secretion.

(5) Of the ten IVSP protein constituents we have analysed, eight are found only in Gram-negative bacteria. Only the two ATPases, VirB4 and VirB11, are also found in Gram-positive bacteria, and only VirB11 is found ubiquitously in a wide variety of bacteria and archaea. In these other organisms, these ATPases clearly do not function as constituents of IVSP systems.

(6) While the structural analyses lead to clear predictions regarding function (see point 3 above), the phylogenetic analyses suggest that IVSP systems have evolved from a single precursor system with virtually no shuffling of constituents between sequence-divergent systems. This important observation is illustrated by the phylogenetic trees shown in Figs 1–4. Moreover, these systems have, for the most part, probably diverged in function without gain or loss of protein constituents. The only exceptions are the proteins of clusters 2, 9 and 10 (see Table 2). The same general conclusion has been reached regarding other macromolecular export systems including the type I secretory pathway (ABC) (Fernandez & de Lorenzo, 2001) and type III secretory pathway (Vir) (Plano et al., 2001) systems (see Kuan et al., 1995; Nguyen et al., 2000; Paulsen et al., 1997). In all three cases, the lack of...
shuffling indicates the occurrence of extensive protein–protein interactions, allowing construction of transenvelope exporters that can secrete their substrates directly from the cell cytoplasm to the extracellular medium without allowing accumulation of a periplasmic intermediate. It is interesting to note that type II secretory pathway systems do allow accumulation of folded periplasmic intermediates, and that export across the two envelope membranes occurs in two distinct steps (Pugsley, 1993; Sandkvist, 2001).

The computational analyses presented in this article provide some confirmatory and some valuable new information about the structures and mechanisms of action of a unique set of protein complexes. The sequence analyses reported should provide guides for detailed molecular genetic analyses. Many of the observations and predictions presented require direct experimental verification. It is hoped that these studies will be forthcoming in the near future.

Acknowledgements

We wish to thank Christian Baron for helpful suggestions and Mary Beth Hiller for her assistance in the preparation of this manuscript. Work in the authors’ laboratory was supported by grants GM 64368-24 and GM 55434-17 as well as by the M. H. Saier, Sr Memorial Research Fund.

References


