Coevolution between nonhomologous but functionally similar proteins and their conserved partners in the Legionella pathogenesis system

Michal Feldman, Tal Zusman, Shelly Hagag, and Gil Segal*

Department of Molecular Microbiology and Biotechnology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat-Aviv, Tel Aviv 69978, Israel

Edited by Stanley Falkow, Stanford University, Stanford, CA, and approved July 5, 2005 (received for review March 7, 2005)

Legionella pneumophila, the causative agent of Legionnaires’ disease, and other pathogenic Legionella species multiply inside protozoa and human macrophages by using the intracellular multiplication (Icm)/defect in organelle trafficking (Dot) type-IV secretion system. The IcmQ protein, which possesses pore-forming activity, and IcmR, which regulates the IcmQ activity, are two essential components of this system. Analysis of the region expected to contain these two genes from 29 Legionella species revealed the presence of a conserved icmQ gene and a large hypervariable gene family [functional homologues of icmR (fir) genes], located at the icmR genomic position. Although hypervariable in their sequence, the fir genes from all 29 Legionella species were found, together with their corresponding icmQ genes, to function similarly during infection. In addition, all FIR proteins we examined were found to interact with their corresponding IcmQ proteins. Detailed bioinformatic, biochemical, and genetic analysis of the interaction between the variable FIR proteins and conserved IcmQ proteins revealed that their interaction depends on a variable region located between two conserved domains of IcmQ. This variable region was also found to be critical for IcmQ self-interaction, and the region probably coevolved with the corresponding FIR protein. A FIR–IcmQ pair was also found in Coxiella burnetii, the only known non-Legionella bacterium that contains an Icm/Dot system, indicating the significance of this protein pair for the function of this type-IV secretion system. We hypothesize that this gene variation, which is probably mediated by positive selection, plays an important role in the evolutionary arms race between the protozoan host cell and the pathogen.

A mong all known Legionella species, Legionella pneumophila is the most common causative agent of Legionnaires’ disease in the world, but several other species were also found to be capable of causing severe pneumonia (1). It is now well established that a type-IV secretion system, which effectors are translocated into the host cell, is the major pathogenesis system of L. pneumophila (2–6). This type-IV secretion system is composed of 25 intracellular multiplication (Icm)/defect in organelle trafficking (Dot) (icm/dot) genes that were shown to be required for the intracellular multiplication of L. pneumophila inside human macrophages and in protozoa (reviewed in ref. 7). Two of the icm/dot genes, icmR and icmQ, are located next to one another, and IcmQ was shown to self-interact and to form pores in lipid membranes, two activities that were inhibited when IcmQ was bound to its chaperone encoded by the icmR gene (8, 9).

Coxiella burnetii, the causative agent of Q fever, was found to contain all the icm/dot genes except icmR (10, 11), and, in two other Legionella species, nonhomologous genes were found instead of the icmR gene (12). These findings prompted us to explore the genetic variation of the icmR–icmQ region and its relation to pathogenesis. To do that, we cloned, sequenced, and analyzed this region from 26 additional Legionella species, covering most of the Legionella evolutionary tree. This analysis revealed a highly variable gene family at the icmR location, but all these genes were found to function similarly, and, thus, they were named, collectively, functional homologues of icmR (fir) genes. A deeper analysis of five Legionella FIR–IcmQ pairs and the C. burnetii FIR–IcmQ pair suggests coevolution between the FIR protein family and their corresponding IcmQ partners.

Materials and Methods

Bacterial Strains, Plasmids, and Media. The Legionella species and strains used in this study are described in Table 1, which is published as supporting information on the PNAS web site. Bacterial media, plates, and antibiotic concentrations were used as described in ref. 13. The plasmids used in this study are listed in Table 2, which is published as supporting information on the PNAS web site.

Cloning of the fir-icmQ Region from 26 Legionella Species. Cloning was performed, as described in ref. 12, for Legionella longbeachae and Legionella micdadei. The fir–icmQ region was amplified, cloned, and sequenced twice from each of the species. The accession nos. of the resulting sequences are listed in Table 1. To construct complementation plasmids, the fir genes, with their corresponding icmQ genes, were cloned into pMMB207aB-Km14 (14), as described in ref. 12.

Plasmid Construction for the Two-Hybrid Analysis. The fir genes and the full-length icmQ genes and their derivatives were amplified by PCR and cloned into Bordetella pertussis cyA two-hybrid vectors (15). IcmQ containing the W26R substitution was generated by sewing PCR (16). The lacZ levels of expression were measured by using the β-galactosidase assay, as described in ref. 11.

Protein Purification and Far Western Analysis. N-terminal GST fusions to five FIR proteins were constructed by amplification of their coding regions and cloning into the pGEX-2T vector (Amersham Pharmacia). Five icmQ genes were fused to a His6 tag at their N termini by PCR amplification and cloning into the pET-15b vector (Novagen) (the plasmids are listed in Table 2). Protein Purification and Far Western Analysis. N-terminal GST fusions to five FIR proteins were constructed by amplification of their coding regions and cloning into the pGEX-2T vector (Amersham Pharmacia). Five icmQ genes were fused to a His6 tag at their N termini by PCR amplification and cloning into the pET-15b vector (Novagen) (the plasmids are listed in Table 2). All the proteins were purified from Escherichia coli BL21 containing the pRep4 plasmid according to the supplier’s instructions. Far Western analysis using these proteins was performed essentially as described in ref. 17.

Intracellular Growth in Acanthamoeba castellanii. Intracellular assays were performed as described in ref. 14.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: Dot, defect in organelle trafficking; FIR, functional homologues of IcmR; Icm, intracellular multiplication.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AY753534, AY753535, and AY860641–AY860664).

*To whom correspondence should be addressed. E-mail: gil@tauex.tau.ac.il.

© 2005 by The National Academy of Sciences of the USA
Results

Recently, it was shown that, in *L. longbeachae* and *L. micdadei*, nonhomologous genes are located in the same genomic position where the *icmR* gene has been shown to be located in *L. pneumophila*. Surprisingly, however, these nonhomologous genes were shown to function similarly to *icmR* and to encode for proteins that interact with the corresponding IcmQ protein (12). These findings raised several questions related to the evolution of the *Legionella* pathogenesis system, such as: (i) how broad is this phenomenon in the *Legionella* genus, (ii) in what way can conserved proteins function with a collection of variable proteins, (iii) what evolutionary forces drive such a phenomenon, and (iv) how do these evolutionary processes serve the pathogenesis system? To answer these questions, the relevant genomic regions of 29 *Legionella* species were analyzed and compared. In addition, detailed bioinformatic, biochemical, and genetic analyses were performed on five of these species, as described below.

Gene Variation Occurs Within the Same Location in 29 Legionella Species. To analyze the *icm/dot* region containing the *icmT, icmS, icmR, icmQ*, and *icmP* genes in *L. pneumophila* from other *Legionella* species, this region was cloned and sequenced from 29 *Legionella* species. Sequence comparison showed that each of the 29 species contained highly conserved *icmT, icmS, icmQ*, and *icmP* genes. However, in the place where the *icmR* gene was expected to be located, each species carried a different gene with no sequence homology to any known gene or protein in the GenBank database. Moreover, these genes encode proteins that are extremely distinct from one another in their primary sequence, their length (71–135 aa), and their predicted pI values (4.59–9.46), three properties that are highly conserved in the other proteins encoded from this region. The genes were named, collectively, *fir* genes, and they are listed and described in Table 3, which is published as supporting information on the PNAS web site. Pairwise sequence comparisons performed on the 29 sequences showed that, whereas the *icmS* and *icmQ* coding sequences are highly conserved, the *fir* coding sequences located between them are completely nonhomologous (Fig. 1). Moreover, the *icmQ* gene was found to be highly conserved, except for a small region (marked with dots in Fig. 1) that seemed to be variable, similar to the *fir* genes. In addition, in the regions between the coding sequences, conserved regulatory elements were observed, a fact that emphasizes the overall conservation of the region at the DNA level, as opposed to the *fir* coding sequences. At the protein level, high conservation among the 29 IcmQ proteins, in comparison with high diversity among the 29 FIR proteins, was observed with radial trees (see Fig. 6, which is published as supporting information on the PNAS web site).

These results can be explained by one of two hypotheses: The diversity of the *fir* genes can result from horizontal gene transfer that occurs into the same genomic location or from rapid incorporation of random mutations into an ancestral gene. To distinguish between these hypotheses, the described DNA sequences were analyzed. It was found that the region containing the *fir* genes does not show any signs of horizontal gene transfer (such as different G + C content or insertion sequence elements). Moreover, when closely related *Legionella* species were analyzed, it was clear that their FIR proteins contain a low degree of sequence similarity (two groups of four FIR proteins and four groups of two FIR proteins were identified, as shown in Fig. 6 and described in Table 3). Importantly, the degree of similarity between these FIR proteins was remarkably lower in comparison with the homology of their neighboring IcmQ proteins (for example, in the comparison between *Legionella shakespeareii* and *Legionella moravica*, their IcmQ proteins were found to be 33% identical and 38% similar, whereas their IcmQ proteins were 85% identical and 94% similar). To further support this observation, it was found that closely related *Legionella* species contain similar evolutionary-tree structure, when determined with both FIR and IcmQ proteins (highlighted in Fig. 6). These results strongly support the argument against horizontal gene transfer and favor the possibility of rapid incorporation of random mutations. This conclusion led us to examine whether positive selection drives the high diversity of the *fir* genes. To examine this notion, we used the program SELECTON (http://selecton.bioinfo.tau.ac.il), which evaluates what selection forces operated on each amino acid: positive selection (which leads to diversity) or purifying selection (which leads to conservation), as determined by the ratio between synonymous and nonsynonymous substitutions that occurred in the gene sequence (18). The examination of the IcmQ proteins showed that, as expected, most of the protein residues went through purifying selection, except for several amino acid positions located in the nonhomologous region of the IcmQ protein that evolved through positive selection. As for the FIR proteins, alignment between these proteins was impossible because of the high diversity in their sequences and lengths. Therefore, we aligned two groups, each containing only four FIR proteins that were found to be related to one another (highlighted in Fig. 6B). The SELECTON results showed that several positions located throughout the whole FIR protein sequence went through positive selection (see Fig. 7, which is published as supporting information on the PNAS web site). The fact that, despite their evolutionary conservation, a specific region in the 29 IcmQ proteins went through positive selection and that a similar process took place at the *fir* genes, might indicate that coevo-
olution occurred between the variable region of IcmQ and each of the FIR proteins.

The FIR Proteins Function Similarly During Intracellular Growth. To examine the function of the FIR–IcmQ pairs, a L. pneumophila icmRQ double mutant was generated, and this strain was used for complementation of intracellular growth in A. castellanii (a natural host of Legionella), with plasmids containing each of the 29 fir genes and their corresponding icmQ genes. The results of this analysis were unambiguous: All fir–icmQ pairs complemented this mutant strain, despite the fact that none of the fir genes contained any sequence homology to the L. pneumophila fir gene icmR (Table 3). Representatives of this analysis are shown in Fig. 2; among them is the only case of partial complementation that was observed with the Legionella geestiana fir gene and its corresponding icmQ gene. This partial complementation fits with the fact that this species is relatively distant from the other Legionella species (Fig. 6A).

Analysis of the Interaction Between the FIR and IcmQ Proteins. To further analyze the FIR and IcmQ proteins, 15 of these protein pairs were examined in a bacterial two-hybrid system for protein interaction. The analysis included at least one representative from each FIR homologous group and several representatives of FIR proteins that have no other FIR homologues, FIR proteins with variable length and different pI values (Table 3). The results of this analysis were very clear: All 15 protein pairs were found to interact with one another. These results fit perfectly with the complementation results, indicating that, despite the enormous sequence variation, all FIR proteins function alike, along with their corresponding IcmQ proteins.

To further analyze the protein interactions, a biochemical approach, showing direct interaction between the examined proteins, was used. Five GST-tagged FIR proteins (IcmR, LigB, MigB, FigA, and HigA) and their corresponding His-tagged IcmQ proteins from L. pneumophila, L. longbeachae, L. micdadei, Legionella feeleii, and Legionella hackeliae were purified and examined in a far Western analysis. In the analysis presented, the five His-tagged IcmQ proteins were fixed on a membrane and overlaid with each of the GST-tagged FIR proteins. It was found that each of the FIR proteins interacted with its natural IcmQ at the highest affinity, but some of the FIR proteins interacted also with other IcmQ proteins (Fig. 3). The two extremes were FigA, which interacted with all five IcmQ proteins, and MigB, which interacted with only its natural IcmQ. The fact that FigA and LigB interacted with the L. pneumophila IcmQ does not necessarily mean that these pairs (FigA or LigB with the L. pneumophila IcmQ) can function together, because complementation analysis of a L. pneumophila icmR mutant with the figA or the ligB genes by themselves did not result in intracellular growth in A. castellanii (data not shown). This result means that the figA and the ligB genes require their corresponding icmQ genes for complementation, indicating that interaction is separable from function in this system, and interaction does not lead directly to function. The reciprocal far Western analysis, in which the GST-tagged FIR proteins were fixed on a membrane, was also performed, and similar results were obtained (data not shown).

Two Regions in the IcmQ Proteins Are Required for Interaction with the FIR Proteins. The results, showing that 15 FIR–IcmQ protein pairs interact with one another (Table 3), brought up the question, how it is possible that the conserved IcmQ proteins interact with such highly diverse FIR proteins? It was reasonable that the variable region of the IcmQ protein is required for this binding (dotted area in Fig. 1). However, the results presented in Fig. 3 suggest a more complicated situation, in which at least some of the FIR proteins are able to bind different IcmQ proteins, perhaps indicating the involvement of a conserved domain in the interaction. To look for the IcmQ domain(s) required for interaction with the FIR proteins, a similarity plot, based on the alignment of the 29 IcmQ proteins, was generated (Fig. 4A). This plot made it possible to divide the IcmQ proteins into three major parts: the N-terminal domain (amino acids 1–44), the C-terminal domain (amino acids 72–200); both are highly conserved, and, between them, lies a region that is highly variable (this region is encoded by the DNA sequence of the icmQ genes, indicated by the dotted area in Fig. 1). There were two possibilities for the function of this variable region; it might function as a linker between the two conserved domains (in this case it will not have any importance for the interaction with the relevant FIR protein), or the variable region might diverge between one IcmQ and another, because it coevolved with its
The N-terminal conserved and the variable regions of IcmQ are required for interaction. (A) Similarity plot of the 29 IcmQ proteins, according to the alignment performed by VECTORNTI. Thin lines below the plot indicate the three constructs, containing different fragments (NVC, NV, and N) of the L. micdadei, L. feeleii, and L. hackeliae IcmQ proteins that were analyzed in the two-hybrid system. (B) The IcmQ constructs were assayed for interaction with the corresponding FIR protein (open bars) or the L. micdadei full-length IcmQ protein (shaded bars). Lm, L. micdadei; Lh, L. hackeliae. The IcmQ fragments are indicated, as in A, as well as two additional constructs containing the N-terminal conserved domain and different parts of the IcmQ variable region (S7 and 65 aa). M.U., Miller units. (C) Two-hybrid analysis of the L. feeleii IcmQ protein with three fragments of the L. feeleii IcmQ protein (LFQ), three fragments of the L. hackeliae IcmQ protein (LHQ), and three fragments of the L. micdadei IcmQ protein (LmQ). (D) Two-hybrid analysis of the five IcmQ proteins containing a substitution of the tryptophan residue located at position 26 to arginine. Each of the five wild-type IcmQ proteins (Q) or the five mutant IcmQ proteins (QW) was assayed for protein interaction with the relevant FIR protein (open bars) or the L. micdadei full-length IcmQ protein (shaded bars). Lp, L. pneumophila; Lt, L. longbeachae; Lm, L. micdadei; Lf, L. feeleii; Lh, L. hackeliae. β-Galactosidase activity was measured in M.U. at stationary phase. The results are the averages (±SDs) of at least three independent clones. (E) Far Western competition analysis, in which a membrane blotted with L. micdadei His-tagged IcmQ protein was incubated with equal amounts of GST-tagged MigB protein (6 μg/ml) and increasing amounts of L. micdadei His-tagged IcmQ (0, 2, 4, 6, 8, and 10 μg/ml, from left to right). The membrane was then probed with an anti-GST antibody to show titration of the interaction between IcmQ and MigB by the L. micdadei IcmQ.

In conclusion, the results presented clearly indicate the importance of the variable region for the interaction. Four additional plasmids were constructed for the two-hybrid analysis, containing the 65 and 57 N-terminal amino acids of the IcmQ protein from L. micdadei and L. hackeliae, respectively. These four constructs were found to interact with MigB and HigA, respectively, to the same extent as the full-length corresponding IcmQ protein (Fig. 4B). These data, showing that the N-terminal 57 amino acids of the L. micdadei and L. hackeliae IcmQ were sufficient for the interaction with the corresponding FIR protein, fit together with previous information regarding the L. pneumophila IcmQ–IcmR protein interaction (9). In addition, the requirement of the L. micdadei IcmQ variable region for the interaction with MigB fits well with the findings showing that the MigB protein interacted with only its natural IcmQ (Fig. 3), implying that the L. micdadei IcmQ protein contains a unique domain that can be recognized only by its specific FIR protein. As for the L. feeleii IcmQ, the fact that its N-terminal conserved domain was sufficient for interaction with FigA also fits the far Western results, showing that FigA interacted with all five IcmQ proteins we examined (Fig. 3), and, therefore, it might be that FigA has evolved to interact with only the N-terminal conserved domain of the IcmQ proteins, and, thus, this domain in the L. feeleii IcmQ is sufficient for the interaction with FigA. If this supposition is correct, the N-terminal conserved domain of the L. micdadei and L. hackeliae IcmQ proteins by themselves should be sufficient for the interaction with FigA, even though they did not interact with their native FIR proteins. The examination of these interspecies interactions revealed that, indeed, FigA was able to interact with the N-terminal conserved domain of the L. micdadei and L. hackeliae IcmQ at the same strength with which FigA interacted with the corresponding full-length IcmQ proteins (Fig. 4C).
importance of the variable region for the interaction between the IcmQ protein and its corresponding FIR partner. However, it seems that the variable region can be separated into two parts, one involved in interaction and the other probably related to the ability of the FIR–IcmQ protein pairs to function together, as indicated by the lack of complementation of the L. pneumophila icmR mutant by the figA or the ligB genes, even though the proteins encoded by these genes interacted with the L. pneumophila IcmQ protein (see above).

In a genetic screen for a L. pneumophila IcmQ loss-of-function mutant, a tryptophane residue at position 26 was identified as essential for the interaction between IcmQ and IcmR (data not shown). This residue is found at the same position in all 29 IcmQ proteins, and it is included in the N-terminal conserved domain (see Fig. 8, which is published as supporting information on the PNAS web site). To determine the importance of the N-terminal conserved domain for the interaction of IcmQ with the FIR proteins, this tryptophane residue (W26R) was substituted in the five IcmQ proteins mentioned above and examined for interaction with the relevant FIR proteins by using the two-hybrid system. The results clearly indicated that this residue is essential for the interaction of all the IcmQ proteins examined, with their FIR pairs (Fig. 4D). The lack of interaction was not because of instability of the mutant IcmQ proteins, because the His-tagged IcmQ mutant was found to be as stable as the wild-type IcmQ protein (see Fig. 9A, which is published as supporting information on the PNAS web site). The lack of interaction (Fig. 4D), which probably leads to lack of intracellular growth complementation (Fig. 9B) by the L. pneumophila W26R IcmQ mutant, points out the high importance of the FIR–IcmQ interaction for pathogenesis.

The *L. micdadei* FIR and IcmQ Proteins Compete for the Same Binding Site on IcmQ. It was previously shown that *L. pneumophila* IcmQ monomers interact with one another to form homopolymers when IcmQ is released from IcmR (8). We wanted to examine whether the IcmQ proteins in other *Legionella* species self-interact via the same domain used for binding of the relevant FIR protein or through another domain. To examine this question, the five *L. micdadei* IcmQ fragments were examined for interaction with the *L. micdadei* wild-type IcmQ. The analysis revealed that the full-length IcmQ protein interaction was the strongest, whereas deletion of the C-terminal conserved domain reduced the interaction, and a further decrease was observed when the complete variable region was deleted (Fig. 4B). Thus, for IcmQ self-interaction, the complete IcmQ protein is required, whereas the interaction with the MigB protein seems to require only the N-terminal conserved domain and part of the variable region and not the C-terminal domain (Fig. 4B). Because the IcmQ N-terminal conserved region seems to be important for both interactions (IcmQ–IcmQ and MigB–IcmQ), the relevance of this domain for the self-interaction of IcmQ was examined by using the *L. micdadei* IcmQ protein containing the W26R substitution. This mutation drastically decreased the ability of the two full-length IcmQ proteins to interact with each other, as was also shown for the interaction with MigB (Fig. 4D). These results suggest that the *L. micdadei* IcmQ protein competes with MigB for binding to the same domain of IcmQ. To test this hypothesis, a modified version of the far Western analysis was used. Equal amounts of the *L. micdadei* His-tagged IcmQ protein were fixed on a membrane and overlaid with a solution containing equal amounts of GST-tagged MigB protein and increasing amounts of the *L. micdadei* His-tagged IcmQ protein. The results obtained were definite. As the amount of the added *L. micdadei* IcmQ increased, the amount of bound MigB decreased (Fig. 4E), indicating that the *L. micdadei* MigB and IcmQ proteins cannot be bound to IcmQ at the same time.

### Table 1

<table>
<thead>
<tr>
<th>ENV-V-T18</th>
<th>NVC-T18</th>
<th>CoxiG-A-T18</th>
</tr>
</thead>
<tbody>
<tr>
<td>T25-ENV</td>
<td>663±35</td>
<td>1540±137</td>
</tr>
<tr>
<td>T25-NVC</td>
<td>950±107</td>
<td>2625±266</td>
</tr>
<tr>
<td>T25-VC</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>T25-ENV</td>
<td>705±114</td>
<td>1372±178</td>
</tr>
<tr>
<td>T25-NVC</td>
<td>698±94</td>
<td>1868±172</td>
</tr>
<tr>
<td>T25-EN</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>T25-N</td>
<td>-</td>
<td>291±48</td>
</tr>
<tr>
<td>T25-CoxigA</td>
<td>787±69</td>
<td>-</td>
</tr>
</tbody>
</table>

**Fig. 5.** Analysis of the interaction between the *C. burnetii* IcmQ and CoxiG-A. The *C. burnetii* IcmQ protein and CoxiG-A were analyzed by using the two-hybrid system. The *C. burnetii* IcmQ protein is schematically illustrated as four boxes. The black box represents the extension domain, the small open box represents the N-terminal conserved domain, the gray box represents the variable region, and the large open box represents the C-terminal conserved domain. The results are presented, in Miller units (M.U.), as average (±SD) of the analysis of at least three independent clones. (−) indicates that no interaction was observed with the pair of proteins analyzed, and the β-galactosidase result was similar to that of the vector control (−80 M.U.). The numbers 18 and 25 represent the two-hybrid fragments used for the construction of the fusions.

*C. burnetii* Contains a Unique FIR–IcmQ Pair. There is just one other known bacterium, *C. burnetii*, in which an almost complete set of genes homologous to the *L. pneumophila icm/dot cluster, missing the icmR gene, has been identified (7). Taking into account the high diversity of the *fir* genes, it seems reasonable that no gene similar to the *Legionella* *fir* genes will be found in *C. burnetii*. However, a functionally similar gene is expected to be present. Moreover, because the *C. burnetii* IcmQ protein contains an extension of 46 aa at its N terminus, the *C. burnetii* FIR protein is expected to be extremely divergent, relative to the other FIR proteins. Several ORFs were identified in the *C. burnetii* icmQ upstream region; however, two-hybrid analysis of these ORFs revealed that none of them interacts with IcmQ (data not shown). Further analysis of this region uncovered a very short ORF (49 aa long) located immediately upstream from the *icmQ* gene, and its stop codon overlaps the first methionine of IcmQ. Two-hybrid analysis revealed that this ORF (named CoxiG-A) strongly interacts with the *C. burnetii* IcmQ (Fig. 5). It is interesting to note that the size of CoxiG-A (49 aa) and the extension of the *C. burnetii* IcmQ (46 aa) fit very nicely with the average length of the *Legionella* FIR proteins (∼100 aa long). This observation directed us to analyze the interaction of CoxiG-A with the *C. burnetii* IcmQ and the interaction between IcmQ monomers. The analysis indicates that CoxiG-A requires the extension domain and the variable region of IcmQ for interaction (Fig. 5), similar to the *L. hackeliae* and *L. micdadei* FIR proteins. On the other hand, the extension domain was found to have a negative effect on the interaction between *C. burnetii* IcmQ monomers. Interactions were the strongest when the extension domain was deleted from both IcmQ monomers (Fig. 5). These results coincide with previous results showing that the *L. pneumophila* IcmR protein can prevent the polymerization of IcmQ monomers (8), which might also be the function of the extension domain of the *C. burnetii* IcmQ. It is interesting to note that, in both *C. burnetii* and *L. micdadei*, IcmQ self-interaction was observed with the N-terminal conserved domain by itself, but the variable region was found to have a critical effect on this interaction (compare Fig. 4B with Fig. 5). The results presented...
about CoxigA and the C. burnetii IcmQ clearly indicate that, even in C. burnetii, a FIR–IcmQ pair exists, and this pair contains both similarities and differences in its function and organization, in comparison with the FIR–IcmQ pairs of the various Legionella species.

Discussion
A pathogenesis-related gene family that is highly variable in its sequence but highly conserved in its function was discovered in 29 different Legionella species. The FIR proteins encoded by these genes were found to function similarly in conjunction with their corresponding IcmQ proteins (as examined for all protein pairs in an intracellular growth experiment), and physical interaction was observed for all FIR–IcmQ pairs examined. The FIR–IcmQ protein interaction was found to depend on the IcmQ N-terminal conserved domain and part of its variable region. In addition, both the IcmQ variable region and the FIR proteins were found to undergo positive selection. Moreover, the fact that a FIR–IcmQ pair was also found in the more distant bacterium C. burnetii might indicate the importance of this protein pair for the function of the Icm/Dot type-IV secretion system. All the above-mentioned characteristics led us to the hypothesis that, in each Legionella species, coevolution occurred between the hypervariable FIR protein and the variable domain of the conserved corresponding IcmQ protein. The relationship between positive selection and pathogenicity has been suggested. In Salmonella enterica, the rfb locus, which encodes enzymes directing the synthesis of the outer surface protein O antigen, goes through positive selection that enables the bacteria to evade protozoan predators (19). According to this information, we assume that coevolution between the IcmQ and its relevant FIR protein occurred to allow the formation of a complex that fits the ecological needs of each Legionella species, possibly enabling survival in various suitable protozoan hosts. Our model predicts that the FIR–IcmQ complex is secreted upon contact with a protozoan host cell; an unknown connection between the host cell and the FIR protein releases the IcmQ protein, allowing its self-interaction; and these IcmQ homopolymers form pores in the host cell membrane. This model fits all the major findings regarding the FIR–IcmQ protein pair: (i) IcmQ forms pores in lipid membranes, and IcmR has been shown to inhibit this function (9). (ii) IcmQ was shown to be exposed on the bacterial cell upon contact with host cells (9). (iii) All the FIR proteins have a similar function, despite their great sequence variability, as shown in this study. (iv) The sequence variation of the FIR proteins and the variable region in the IcmQ protein coevolved through positive selection, as shown in this study.

Our model also predicts the forces that might drive the variability of the FIR proteins. Looking at the model from the point of view of the protozoan cells, a protozoan cell that will no longer be recognized by a FIR protein will not be used by Legionella as a host, because such a protozoan host cell will not lead to the displacement of the IcmQ protein from the FIR protein because of binding of FIR to a host factor. From the bacterial point of view, FIR proteins that will recognize a large number of protozoan host cells will give an advantage to the Legionella species harboring them. These two forces, the selection for protozoan cells that can avoid Legionella infection and the selection for Legionella species that can infect a broad host range of protozoan cells, might drive this hypervariability of the FIR proteins.

Our data show evidence concerning the intimate relation among evolution, interaction, and function of virulence proteins, information that might deepen the understanding of how the Legionella pathogenesis system evolves and adapts in nature.

We thank Karen Pomeranz for the random mutagenesis screen, Dr. Tal Pupko and Adi Doron-Faigenboim for helpful assistance in using the selection program, and Prof. Martin Kupiec for carefully reading the manuscript. This work was supported by a grant from The Center for the Study of Emerging Diseases and (in part) by Grant 5748 from the Chief Scientist’s Office of the Ministry of Health, Israel (to G.S.).