Supporting Information

Hsu et al. 10.1073/pnas.1207903109

SI Materials and Methods

Bioinformatic Search. The protein sequences of the *Legionella pneumophila* strain Philadelphia 1 were downloaded from NCBI. A Perl script (see below) was constructed to search the compiled sequence file for proteins that contain the CX_5R signature motif. The proteins that contain the CX_5R motif and previously reported to be T4SS substrates were listed in Table S1.

```
#!/usr/bin/perl -w
open(INP, "oneline.txt"); # the original input
file open(OUT, ">of interest.txt"); # output
@a=<INP>;
close INP;
for ($r=0; $r<@a; $r+=2){
chomp $a[$r] ;
if ($a[$r]=~/C.....R/){
print OUT $a[$r-1] , $a[$r] , "\n";
}
}
close OUT;
```

Crystallization. Crystals were grown at room temperature by the hanging-drop vapor diffusion method by mixing 1 μL of protein (10 mg/mL) with an equal volume of reservoir solution containing 0.1 M Tris·HCl at pH 8.0, 10% (wt/vol) PEG 3350, and 10% (vol/vol) ethylene glycol. Rod-shaped crystals were formed within 2–3 d. For phase determination, protein crystals were soaked in cryoprotectant [0.1 M Tris·HCl at pH 8.0, 10% (wt/vol) PEG 3350, and 25% (vol/vol) ethylene glycol] with the addition of 10 mM ethylmercury chloride (kind gift from Dr. Steve Ealick at Cornell University) for 2 h at room temperature. For SidF–substrate complex crystal formation, purified SidF C645S mutant proteins (10 mg/mL) was mixed with a final concentration of 0.3 mM diC4-PI(3,4)P_2 (Echelon Biosciences). The protein–substrate complex was crystallized in a similar condition as described above.

Data Collection and Processing. Diffraction datasets for native and mercury derivative crystals were collected at the Cornell synchrotron light source, MacCHESS beam line A1. The dataset for SidF–substrate complex crystal was collected at Brookhaven National Laboratory, X4C beamline. All datasets were indexed, integrated, and scaled with HKL-2000 (1). The crystals belong to space group $P2_12_12_1$ with a = 71.23 Å; b = 114.83 Å; c = 124.37 Å; $\alpha = \beta = \gamma = 90^0$ (Table S1). The calculated Matthews coefficient V_m= 3.03, with 59.3% of solvent in the crystal and one protein molecule in an asymmetric unit (2).

Structure Determination and Refinement. Three mercury sites corresponding to residues C206, C635, and C602 were identified in the crystal by using the program HKL2MAP (3). The initial phase was calculated by single isomorphous replacement with anomalous scattering (SIRAS) method and was improved by solvent flattening in HKL2MAP. The ab initio protein model was then built with COOT (4). Iterative cycles of model building and refinement were carried out with the with refmac5 (5) in the CCP4 suite (6) to complete the final model.

Enzymatic Assays. All diC8-phosphoinositides were purchased from Cell Signals. All reactions were performed in a polystyrene 96-well plate for 20 min at 37 °C with a total volume of 50 μL of reaction mixture, which contains reaction buffer (50 mM Tris·HCl at pH 8.0 and 150 mM NaCl), 1 nmol of lipids, and 0.1 μg of purified enzymes. Phosphate release was measured at OD_{620} absorbance with the addition of malachite green reagent as described by Maehama et al. (7).

Thin-Layered Chromatography. The green fluorophore, Bodipy-FL labeled diC6 phosphoinositides were purchased from Echelon Research Laboratories. Most reactions were conducted in 20 μL of buffer containing 50 mM ammonium carbonate at pH 8.0 and 2 mM DTT, and with 1 μg of lipid substrates and 0.1 μg of purified proteins for 20 min at 37 °C. For reactions containing two enzymes, the second enzyme was added after the completion of the first SidF catalyzed reaction and was incubated for additional 20 min. Reaction products were dried in a Speed-Vac for 30 min at 45 °C. The dried pellets were resuspended in 10 μL of methanol/isopropanol/acetic acid (5/5/2) and spotted onto a TLC Silica gel 60 F_{254} (EMD) that was pretreated by soaking in methanol/water (3:2) containing 1% potassium oxalate and then 1 h drying in a 65 °C oven. The TLC plates were developed in a solvent system consisting of 1-propanol/2 M acetic acid (65%:35%). Fluorescent lipids were visualized under UV lights by using a Bio-Rad Gel dock system.

Cell Culture and Transfection and Fluorescence Microscopy. Cos7 cells were maintained and grown in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% FBS (Cellgro) and penicillin-streptomycin solution (Cellgro). Cells were transfected with 0.5 μg of each plasmid by using polyethylenimine reagent at 1:5 ratio. Overnight transfected cells were fixed with 4% formaldehyde. Fluorescent microscopy images were acquired with a spinning disk confocal microscopy system (3i) by using a DMI 6000B microscope (Leica) and a digital camera (QuantEM; Photometrics).

Bacterial Infection. The *L. pneumophila* Philadelphia-1 wild-type strain Lp02, the type IV secretion system defective strain Lp03 (dotA-), the *sidF* deletion mutant, and the complementation strain were described (8). *L. pneumophila* was grown and maintained on CYE medium as described (9). For infection, bacterial strains were grown to the postexponential phase as determined by optical density of the cultures (OD_{600} = 3.3–3.8) as well as an increase of bacterial motility.

Immunostaining. Mouse bone marrow-derived macrophages were infected with relevant strains for indicated time points. Samples were fixed by following standard protocols. SidF and SidC were stained by using specific antibodies described in earlier studies (8, 10).

1. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
2. Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* 33:491–497.
3. Pape T, Schneider TR (2004) HKL2MAP: A graphical user interface for macromolecular phasing with SHELX programs. *J Appl Cryst* 37:843–844.
4. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
5. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53: 240–255.
6. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50:760–763.
7. Maehama T, Taylor GS, Slama JT, Dixon JE (2000) A sensitive assay for phosphoinositide phosphatases. *Anal Biochem* 279:248–250.

8. Banga S, et al. (2007) Legionella pneumophila inhibits macrophage apoptosis by targeting pro-death members of the Bcl2 protein family. *Proc Natl Acad Sci USA* 104:5121–5126.
9. Conover GM, Derré I, Vogel JP, Isberg RR (2003) The Legionella pneumophila LidA protein: A translocated substrate of the Dot/Icm system associated with maintenance of bacterial integrity. *Mol Microbiol* 48:305–321.
10. Luo ZQ, Isberg RR (2004) Multiple substrates of the Legionella pneumophila Dot/Icm system identified by interbacterial protein transfer. *Proc Natl Acad Sci USA* 101: 841–846.

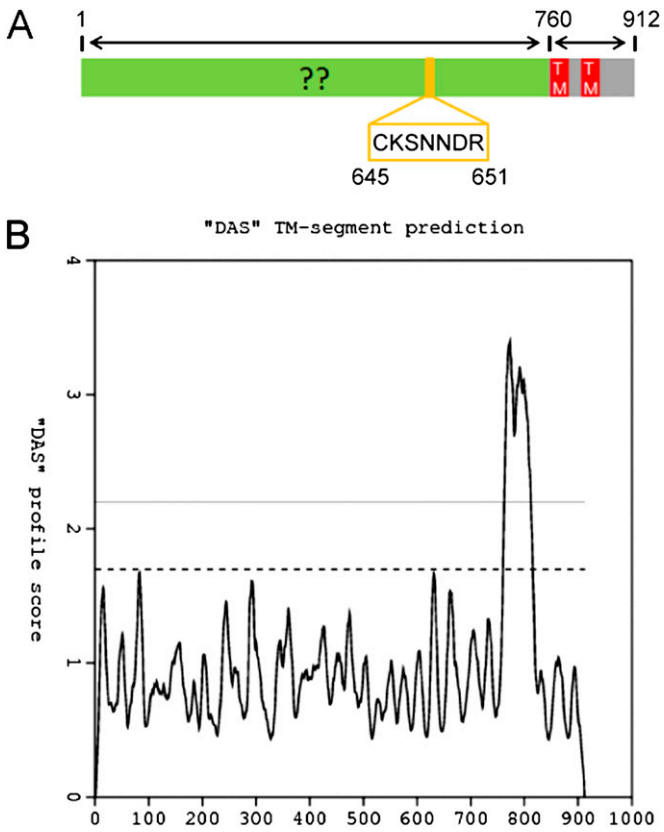

Fig. S1. Schematic domain structure of SidF. (*A*) SidF is a protein comprised of 912 residues. It contains a "CX_5R" signature motif that is characteristic of PI phosphatases and protein tyrosine phosphatases. (*B*) The C terminus of SidF contains two transmembrane motifs predicted by the DAS Transmembrane Prediction server (1).

1. Cserzö M, Wallin E, Simon I, von Heijne G, Elofsson A (1997) Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: The dense alignment surface method. *Protein Eng* 10:673–676.

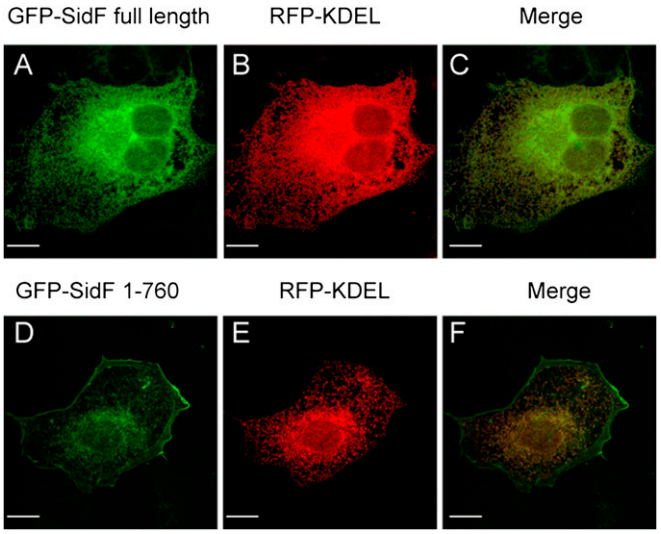

Fig. S2. Intracellular localization of transfected SidF in mammalian cells. (*A–C*), Exogenously overexpressed EGFP-SidF colocalizes with ER marker, RFP-KDEL, in Cos7 cells. (*D–F*), The cytosolic portion (1–760) of SidF localizes to the plasma membrane when overexpressed in Cos7 cells. (Scale bars: 10 μm.)

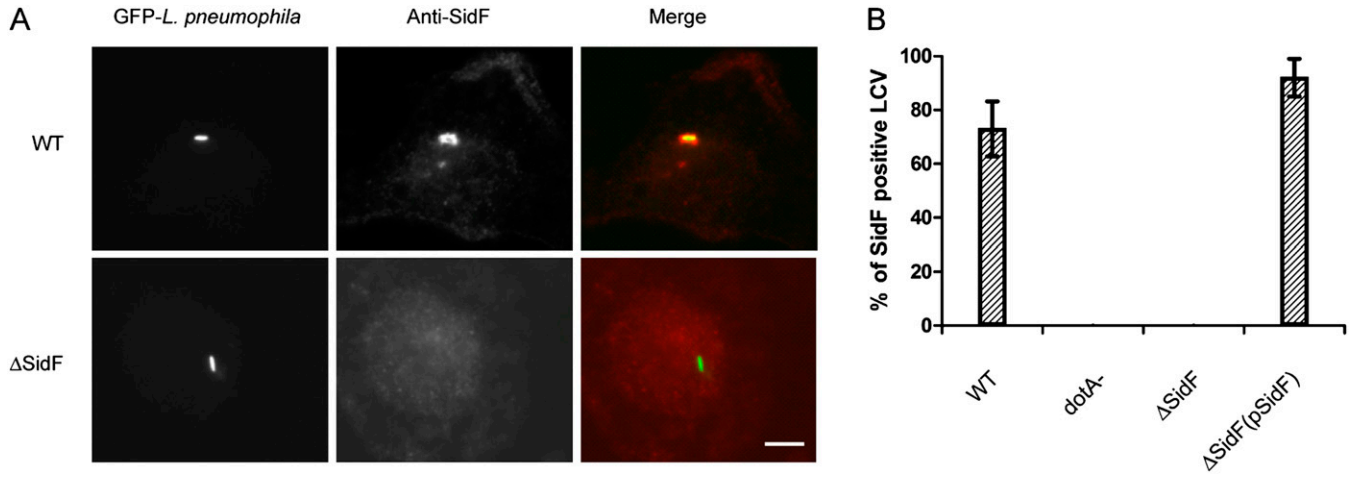

Fig. S3. Association of SidF with the *Legionella*-containing vacuole. (*A*) Mouse bone marrow-derived macrophages were infected with *L. pneumophila* strains expressing GFP grown at an MOI of 5 for 2 h. Infected samples were fixed and stained for extracellular bacteria with rat anti-*Legionella* antibody (green). SidF was detected using a specific antibody (8) and a secondary antibody conjugated to Texas red. SidF is enriched on the surface of LCV membrane. (Scale bars: 5 μm.) (*B*) Quantification of SidF-positive LCVs. Data shown are from two independent experiments performed in triplicate in which at least 100 phagosomes were scored per coverslip. dotA-, the type 4 secretion system defective strain Lp03(dotA-); ΔSidF, the *sidF* deletion mutant Lp02 strain; and ΔSidF(pSidF), the *sidF* deletion Lp02 strain complemented with a plasmid expressing SidF; WT, *L. pneumophila* Philadelphia-1 wild-type strain Lp02.

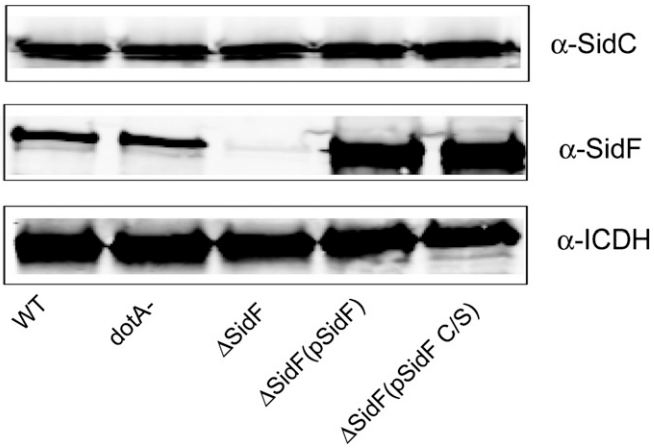

Fig. S4. Expression of SidC and SidF in *L. pneumophila* strains. Lysates of the strains used for infection were separated by SDS/PAGE and probed for SidC (*Top*) and SidF (*Middle*). The isocitrate dehydrogenase (ICDH) was detected as a loading control (*Lower*). dotA-, the type IV secretion system defective strain Lp03 (dotA-); ΔSidF, the *sidF* deletion mutant Lp02 strain; ΔSidF(pSidF), the *sidF* deletion Lp02 strain complemented with a plasmid expressing SidF; ΔSidF (pSidF C/S), the *sidF* deletion Lp02 strain complemented with a plasmid expressing SidF C/S catalytic dead mutant; WT, *L. pneumophila* Philadelphia-1 wild type strain Lp02.

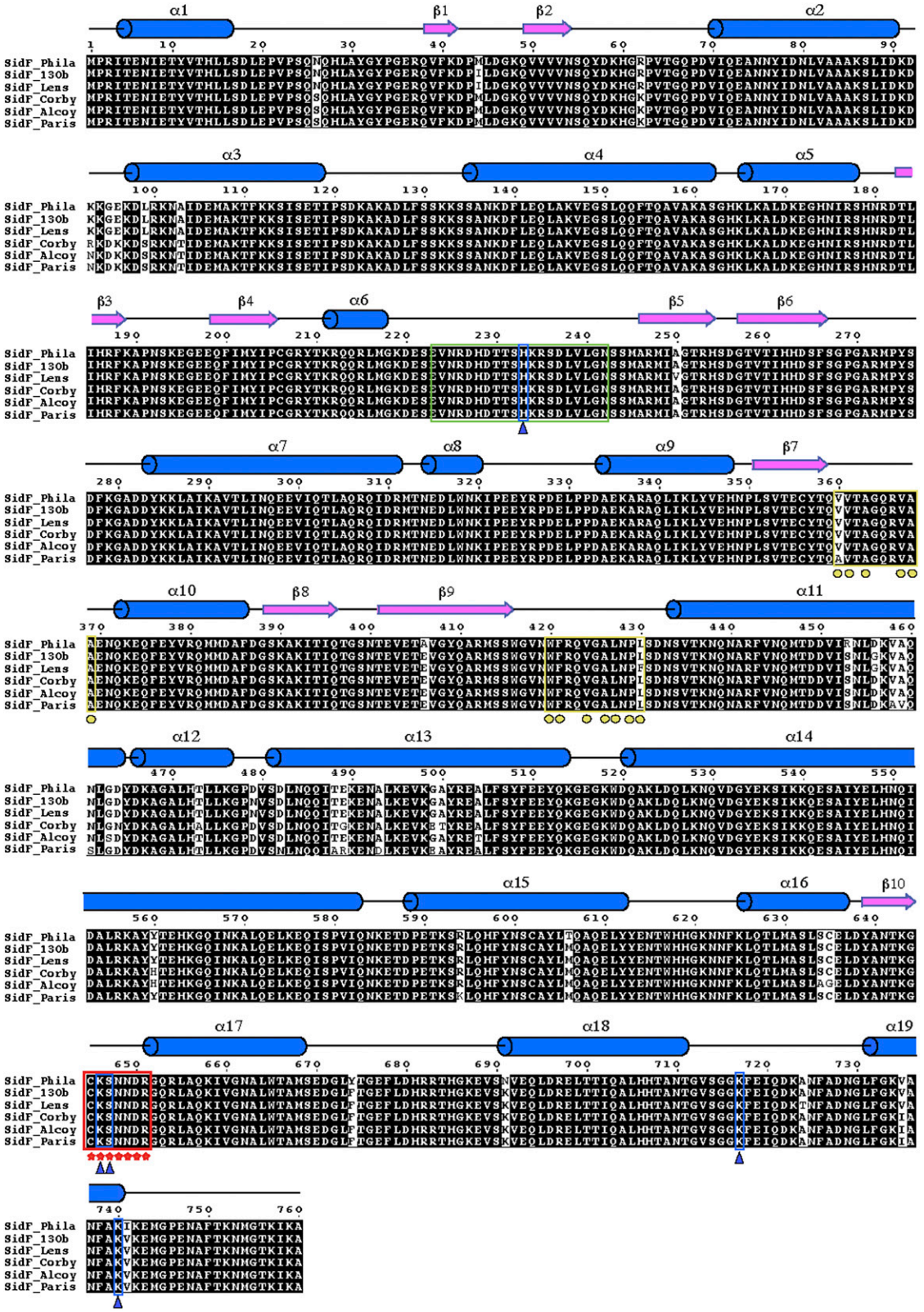

Fig. S5. Multiple sequence alignment of SidF. The sequences corresponding to the SidF 1–760 region from different *Legionella* species were aligned by MutiAlin (1) and colored by ESPript 2.2 (2). Secondary elements are drawn above the alignment. The catalytic residues are marked with red stars. Residues important for D4 phosphate recognition are highlighted with blue triangles. Two hydrophobic loops are marked with gold circles (colored gold in Fig. 2 *A*–*C*). The flexible loop containing His233 is marked with a green box. Entrez database accession numbers are as follows: SidF_Phila, gi: 52842790; SidF_130b, gi: 307611464; SidF_Lens, gi: 54295421; SidF_Corby, gi: 148255306; SidF_Alcoy, gi: 296108226; and SidF_Paris, gi: 54298573.

1. Corpet F (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 16:10881–10890.
2. Gouet P, Courcelle E, Stuart DI, Métoz F (1999) ESPript: Analysis of multiple sequence alignments in PostScript. *Bioinformatics* 15:305–308.

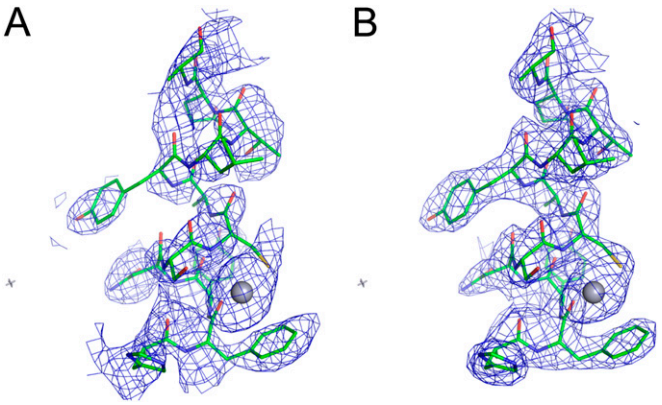

Fig. S6. Representative electron density maps. (*A*) Electron density map contoured at 1σ after SIRAS phasing. (*B*) Electron density at the same area after refinement. The mercury atom is shown in gray sphere.

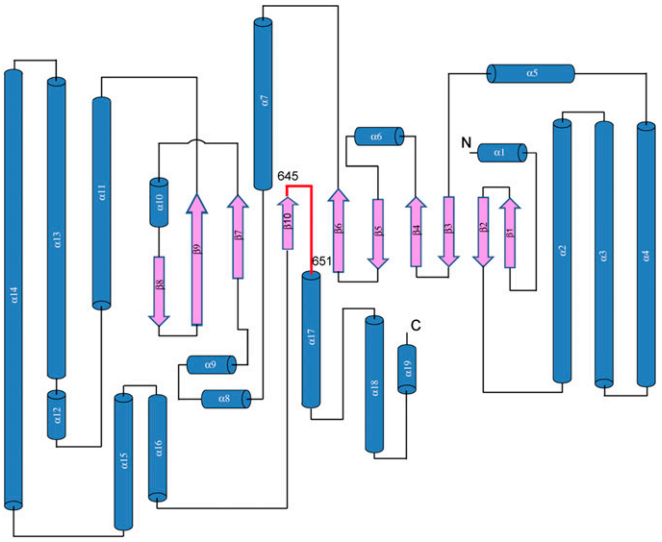

Fig. S7. Schematic diagram of the secondary structure topologies of SidF. Alpha helixes are represented by cylinders and β strands are represented by arrows. The catalytic "CX_5R" motif is numbered and highlighted in red.

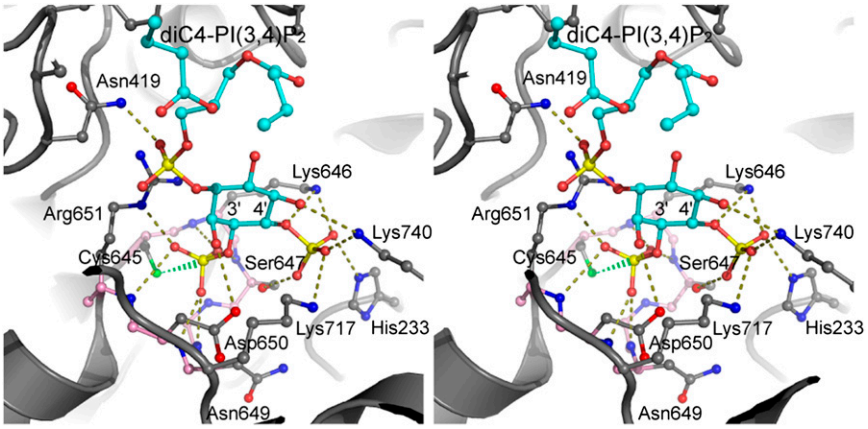

Fig. S8. Stereoview of the bound diC4-PI(3,4)P_2 substrate at the catalytic site. Residues that are involved in hydrogen bonding with diC4-PI(3,4)P_2 molecule are shown in sticks. Hydrogen bonds are indicated by dashed lines between donor and acceptor. The "CX_5R" motif is colored in light pink. The D3 phosphate is hydrogen bonded with main chain amide groups of Lys646, Ser647, Asn649, Asp650, and Arg651 and the guanidinium group of Arg651. The D4 phosphate is engaged in a cationic pocket as described in Fig. 4.

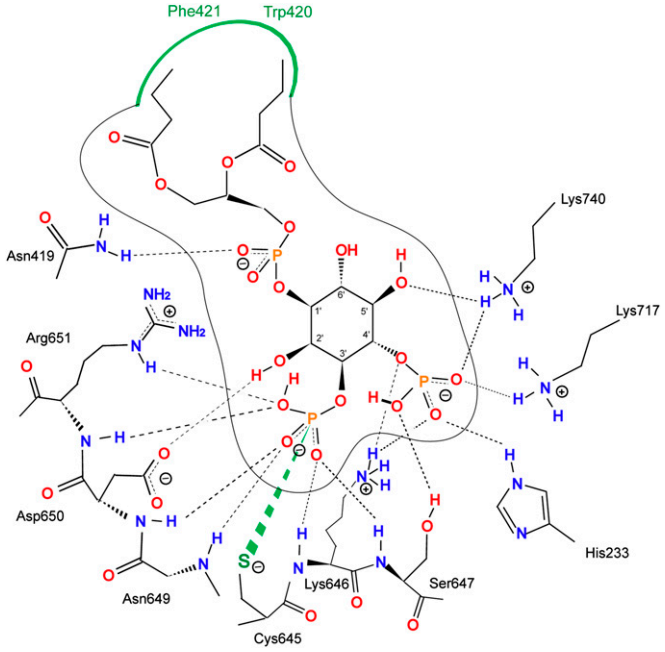

Fig. S9. Two-dimension depiction of SidF–diC4-PI(3,4)P_2 complex. This schematic diagram is modified from the calculated results by using the PoseView server (1). Hydrogen bonding is represented by dashed lines, and hydrophobic interaction between diC4-PI(3,4)P_2 and Trp420 and Phe421 is indicated by a green curve.

1. Stierand K, Maass PC, Rarey M (2006) Molecular complexes at a glance: Automated generation of two-dimensional complex diagrams. *Bioinformatics* 22:1710–1716.

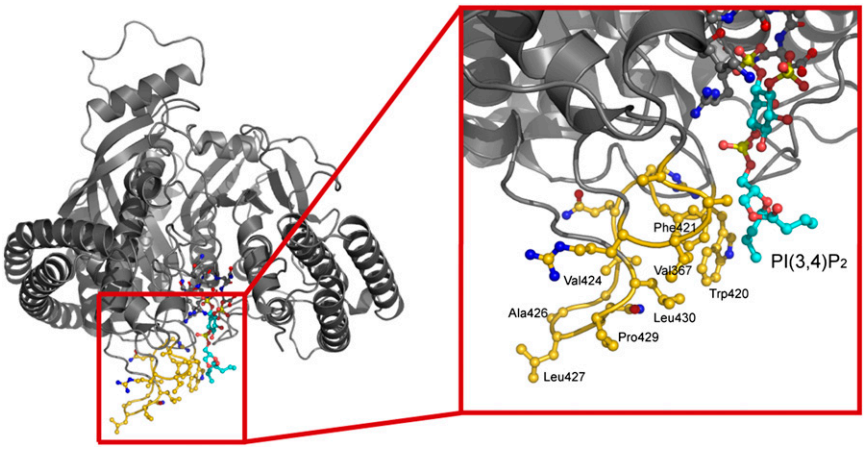

Fig. S10. Two lipids interacting hydrophobic loops in SidF. Two loops enriched with hydrophobic residues protrude out from a flat surface containing the catalytic site. Hydrophobic residues within these two loops are shown in sticks and colored in gold. The substrate diC4-PI(3,4)P_2 is shown in sticks with its carbon atoms colored in cyan. Note that the hydrophobic lipid tails of the substrate form hydrophobic interactions with Trp420 and Phe421.

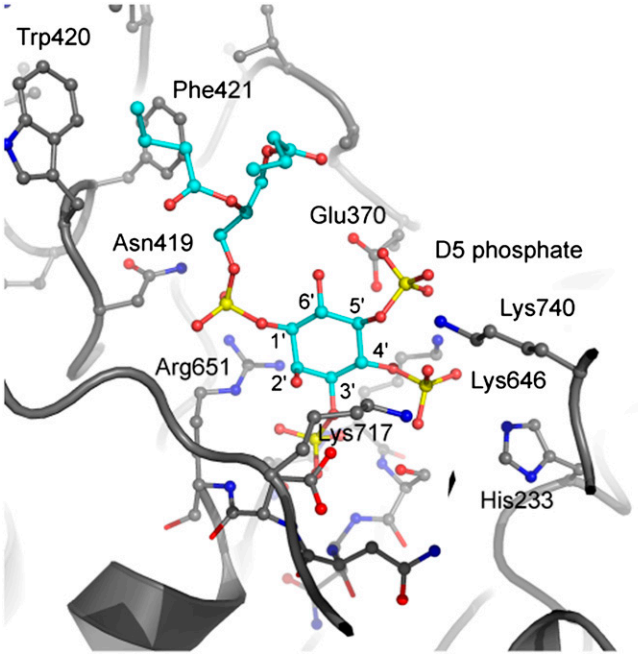

Fig. S11. Structure model of SidF bound with diC4-PI(3,4,5)P_3. The diC4-PI(3,4,5)P_3 molecule is placed at the catalytic site with the D3 and D4 phosphates having the same positions as in diC4-PI(3,4)P_2. In this model, the D5 phosphate is facing outward. Glu370, which is close to the D5 position, may repel the D5 phosphate group. This model may explain our observed higher in vitro enzymatic activity of SidF against diC8-PI(3,4)P_2 than that against diC8-PI(3,4,5)P_3.

Table S1. Effector proteins from *Legionella pneumophila* strain Philadelphia 1 that contain the "CX_5R" motif

Gene ID	Description	Length, aa
lpg0012	Hypothetical protein	525
lpg0090	Hypothetical protein	1324
lpg0130	Hypothetical protein	822
lpg0284	Hypothetical protein	374
lpg0285	Hypothetical protein	227
lpg0365	Hypothetical protein	896
lpg0963	Hypothetical protein	413
lpg0967	Hypothetical protein	197
lpg1120	Hypothetical protein	588
lpg1145	Hypothetical protein	769
lpg1183	Hypothetical protein	630
lpg1290	Hypothetical protein	528
lpg1449	Hypothetical protein	870
lpg1453	Hypothetical protein	172
lpg1484	Hypothetical protein	269
lpg1683	Hypothetical protein	502
lpg1689	23.7 kda protein	208
lpg1751	Hypothetical protein	436
lpg1924	Hypothetical protein	930
lpg1959	Hypothetical protein	664
lpg1963	Hypothetical protein	699
lpg2525	Hypothetical protein	444
lpg2526	Hypothetical protein	455
lpg2529	Hypothetical protein	572
lpg2541	Hypothetical protein	277
lpg2584	SidF	912
lpg2628	Membrane protein	250
lpg2718	Hypothetical protein	520
lpg2912	Hypothetical protein	495
lpg1683	Hypothetical protein	502

Table S2. Data collection, phasing, and structural refinement statistics

	Native	Hg	diC4-PI(3,4)P_2
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions, Å	a = 72.23, b = 115.96, c = 124.22, α = β = γ = 90°	a = 71.84, b = 115.71, c = 125.11, α = β = γ = 90°	a = 71.23, b = 114.83, c = 124.37, α = β = γ = 90°
Data collection statistics			
Synchrotron beam lines	MacHESS A1	MacHESS A1	NSLS X4C
Wavelength, Å	0.9789	0.9789	0.9789
Maximum resolution, Å	2.40	2.42	2.82
Observed reflections	601,286	582,094	301,119
Unique reflections	40,455	75,517	24,387
Completeness, %	99.9 (99.9)	99.6 (99.5)	96.8 (78.9)
$\langle I\rangle/\langle\sigma\rangle$*	27.8 (3.89)	22.3 (2.8)	28.6 (3.0)
R_{sym}*†, %	6.9 (49.8)	11.0 (57.6)	10.2 (44.1)
No. of Hg sites		3	
Refinement statistics			
Resolution*, Å	84–2.41 (2.49–2.41)	47–2.42 (2.49–2.42)	36–2.82 (2.9–2.82)
R_{crys}/R_{free}*‡, %	21.0/25.6 (33.2/35.9)	21.1/26.9 (29.6/36.1)	18.3/25.2 (27.4/34.6)
rms bond length, Å	0.017	0.017	0.012
rms bond angles, °	2.067	2.029	1.66
Mean B factors, $Å^2$			
Protein	43.94	39.68	45.16
Water	53.92	56.67	53.61
diC4-PI(3,4)P_2			83.42
Ramachandran plot			
Most favored/Additional, %	95.8/4.2	96.4/3.6	94.3/5.7
Generous/Disallowed, %	0/0	0/0	0/0

*Values in parentheses are for the highest resolution shell.

†$R_{sym} = \sum_h\sum_i|I_i(h) - \langle I(h)\rangle|/\sum_h\sum_i I_i(h)$.

‡$R_{crys} = \sum(|F_{obs}| - k|F_{cal}|)/\sum|F_{obs}|$. R_{free} was calculated for 5% of reflections randomly excluded from the refinement.