

The solution structure of the BPSL1445 protein of *Burkholderia pseudomallei* reveals the SYLF domain three-dimensional fold

Giacomo Quilici¹, Andrea Berardi¹, Chantal Fabris¹, Michela Ghitti¹, Marco Punta², Louise J Gourlay^{3,4}, Martino Bolognesi^{3,4}, Giovanna Musco¹

¹Biomolecular NMR Laboratory, I.R.C.S.S Ospedale San Raffaele, Via Olgettina 58, 20132 Milan (Italy)

²Center for Omics Sciences, I.R.C.C.S. Ospedale San Raffaele, Via Olgettina 58, 20132 Milan (Italy)

³Department of Biosciences, University of Milano, Via Celoria 26, 20133 Milan (Italy)

⁴Centro di Ricerca Pediatrica Romeo ed Enrica Invernizzi, Università degli Studi di Milano, 20133 Milan (Italy)

Abstract

The SYLF domain is an evolutionary conserved protein domain with phosphatidylinositol binding ability, whose three-dimensional structure is unknown. Here we present the solution structure and the dynamics characterization of the SYLF domain of the bacterial BPSL1445 protein. BPSL1445 is a seroreactive antigen and a diagnostic marker of *Burkholderia pseudomallei*, the etiological agent of melioidosis, a severe infectious disease in the tropics. The BPSL1445 SYLF domain (BPSL1445-SYLF) consists of a β -barrel core, with two flexible loops protruding out of the barrel and three helices packing on its surface. Our structure allows for a more precise definition of the boundaries of the SYLF domain compared to previous knowledge and suggests common ancestry with the bacterial EipA domain. We also demonstrate by PiP arrays and NMR titrations that BPSL1445-SYLF weakly interacts with phosphoinositides, thus supporting lipid binding abilities of this domain also in prokaryotes.

Introduction

The SYLF domain is an evolutionary conserved domain, whose name derives from its representative protein members SH3YL1, Ysc84p/Lsb4p, Lsb3p and plant FYVE. It is also termed YAB (Ysc84 Actin Binding) domain because it confers actin binding ability to the yeast protein Ysc84^{1,2}. In the human SH3YL1 (SH3 domain-containing YSC84-like protein 1) protein, where it is responsible for the regulation of dorsal ruffle, it acts as phosphoinositide-interacting module³. Its lipid binding properties have been further confirmed in other studies performed in yeast Ysc84² and in the *Arabidopsis thaliana* CSF1 protein⁴. While the structures of several phosphoinositide binding domains such as PH (pleckstrin homology), ENTH (epsin N-terminal homology), CALM (clathrin assembly lymphoid myeloid), ANTH (AP180 N-terminal homology), PTB (phosphotyrosine binding), BAR (Bin, amphiphysin, and Rvs), PX (phox homology), postsynaptic density protein, Drosophila disc large tumor suppressor, PDZ (Zonula occludens-1 protein), FERM (F-actin binding ezrin, moesin, and radixin), FYVE (Fab 1, YOTB, Vac 1, and EEA), C2 and Tubby domain (reviewed in ⁵⁻⁷) are well described, the three-dimensional structure of the SYLF domain is still unknown, and the molecular details of its phospholipid binding activity are still elusive. To gain structural and functional insights into this poorly characterized domain, we have solved the structure of the SYLF domain of BPSL1445 (BPSL1445-SYLF). BPSL1445 is a serodiagnostic protein of *Burkholderia pseudomallei*, a pathogenic, aerobic, soil-dwelling Gram-negative bacterium responsible for melioidosis. This is a severe endemic disease, ubiquitous in the tropics^{8,9} with a high fatality rate (10–50%)^{10,11} and clinical presentations ranging from abscess formation to pneumonia and sepsis⁹. Due to its airborne transmission, high mortality, multi-drug resistance and absence of both efficient diagnostic methods and a protective vaccine, *B. pseudomallei* is listed as potential biological warfare and biological terrorism agent⁹. BPSL1445 is a highly reactive antigen against immune sera from *B. pseudomallei* infected patients¹² and is thus considered as a diagnostic signature protein¹³. As such, within the framework of a structural vaccinology project¹⁴⁻¹⁸, BPSL1445 was also indicated as a potential candidate for the discovery of new antigens/epitopes for the development of a melioidosis vaccine and/or of diagnostics¹³. BPSL1445 is conserved in all strains of *B. pseudomallei*, but very little information on its function is available: it is annotated as putative lipoprotein, with documented high expression during the early log phase of growth¹⁹ and up-regulation upon exposure of the bacterium to human serum or insulin²⁰. Over-expression of BPSL1445 has also been observed in a hamster model of melioidosis²¹,

confirming its potential diagnostic relevance for the disease. Despite the current limited knowledge of this seroreactive antigen, its possible clinical relevance makes it an ideal target on which to study the structural and functional characteristics of SYLF domains. In this work we offer the first molecular insights into the SYLF domain fold, and show that it consists of a β -barrel core, with two flexible loops protruding out of the barrel and three helices that pack on its surface. Moreover, we demonstrate by lipid-protein interaction assays and nuclear magnetic resonance (NMR) titrations that BPSL1445-SYLF weakly interacts with phosphoinositides, thus supporting the lipid binding propensity of this domain.

Results and Discussion

Three-Dimensional Structure of BPSL1445-SYLF domain.

Preliminary amino acid sequence analysis of BPSL1445 highlighted the presence of two distinct regions within BPSL1445 primary structure. The first is composed of an N-terminal signal peptide (residues 1-34) harboring a consensus lipobox LxxC (x, any amino acid) motif (L₂₁AGC₂₄)²² with a putative cysteine lipidation site and two predicted cleavage sites at position G23 and A34, respectively (Supplementary Figure S1A). The second region, about 160 residues in length and covering the rest of the protein, is annotated in several protein family databases (CDD²³, Pfam²⁴, InterPro²⁵) as belonging to the SYLF (or Ysc84) family. The SYLF domain is conserved from bacteria to mammals (Figure 1A). Intrigued by the presence of a SYLF domain within the BPSL1445 sequence and aiming to provide insight into the structural-functional relationship of this poorly explored domain, we expressed and purified a construct comprising residues A35 to I195 of BPSL1445 and determined its solution structure by multidimensional heteronuclear NMR spectroscopy (Table 1). The good peak dispersion of the ¹H-¹⁵N-HSQC (heteronuclear single quantum coherence) spectrum²⁶, the far-UV Circular Dichroism (CD) spectrum and the high melting temperature (T_m=77°C), indicated that the construct is stable and well-folded (Figure 1B,C). The domain is monomeric in solution, as assessed by its elution volume in size exclusion chromatography, that is compatible with a protein domain of 17kDa (Supplementary Figure S2A,B). The structural core of BPSL1445-SYLF consists of eight twisted antiparallel β -strands and two well-ordered loops, L3 (V152-P168) and L4 (T175-T188), that together form a β -barrel of \approx 15Å length and \approx 10Å diameter. Two flexible loops L1 (V78-G90) and L2 (L109-S119) link β 1 with β 2 and β 3 with β 4, respectively, and a well-ordered double turn (DT) connects β 5 with β 6. Two amphipathic α -

helices ($\alpha 1$ and $\alpha 2$), consisting respectively of a long N-terminal α -helix (N37-T58) and a two turn α -helix (S62-K68) are connected by a three-residue turn (T1) and are arranged in a L-shaped fold, that packs on one side of the β -barrel, herewith creating extensive hydrophobic contacts.

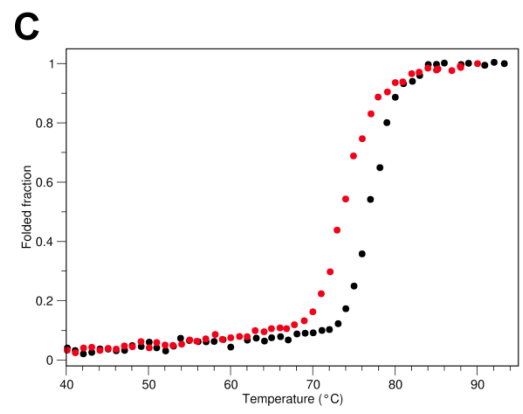
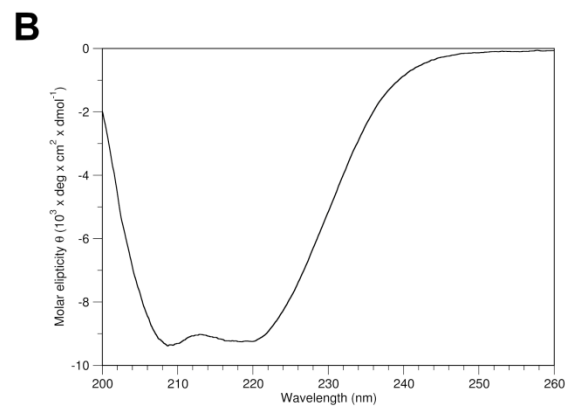
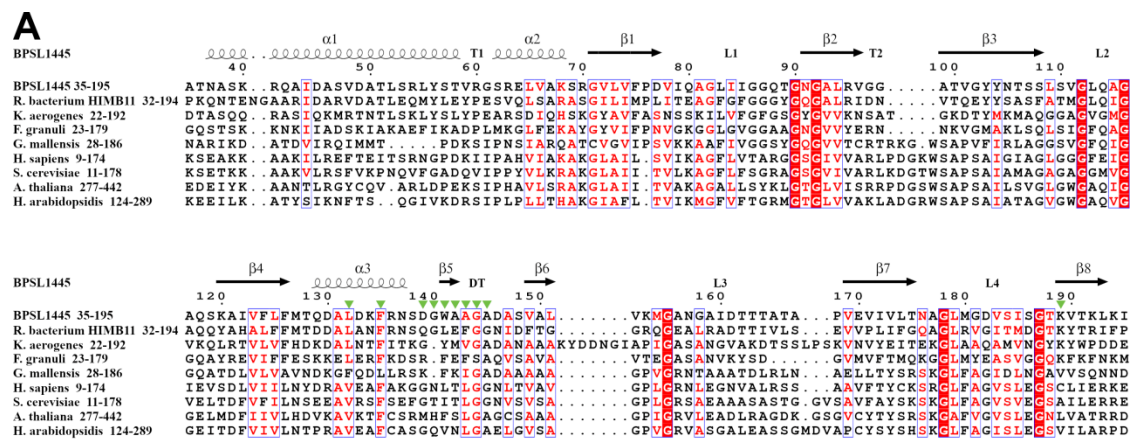


Figure 1: Sequence alignment and Circular Dichroism (CD) of SYLF-domain (A) Multiple sequence alignment of the SYLF domain of representative sequences from different taxonomic groups. The elements of secondary structure present in BPSL1445-SYLF 3D structure are indicated on the top (ESPrpt)⁵² (identical residues, white on red background; partially conserved residues, boxed and red). Green triangles indicate residues significantly affected by PiPs binding in ¹H-¹⁵N HSQC spectra. **(B)** Far-UV CD spectrum at 20°C of BPSL1445-SYLF showing the typical α -helical profile, with a local double minimum at 208 and 220 nm. **(C)** BPSL1445-SYLF CD thermal denaturation curve (expressed as fraction of folded protein) monitored at 220 nm without (black dots) and with (red dots) a 20-fold excess of Pi3P.

The top of the β -barrel is closed by a third helix ($\alpha 3$) (Q129-N137), whereas the bottom is partially obstructed by loops L1 and L2 (Figure 1A, Figure 2A,B). No cavities or empty spaces are found within the barrel, the lumen of which is filled by hydrophobic branched amino-acids and an aromatic core (F75, Y104, F124, F135 and W141) that contribute to overall fold stability (Figure 2C). The C-terminal tail, that is in close proximity to the N-terminus, is well-ordered, with L193 and I195 establishing an extensive network of Nuclear Overhauser effect (NOE)

interactions with Y103 (on β 3) I44 (on α 1) (Figure 2C). The overall topology of BPSL1445-SYLF is α 1-T1- α 2- β 1-L1- β 2-T2- β 3-L2- β 4- α 3- β 5-DT- β 6-L3- β 7-L4- β 8 (Figure 2D).

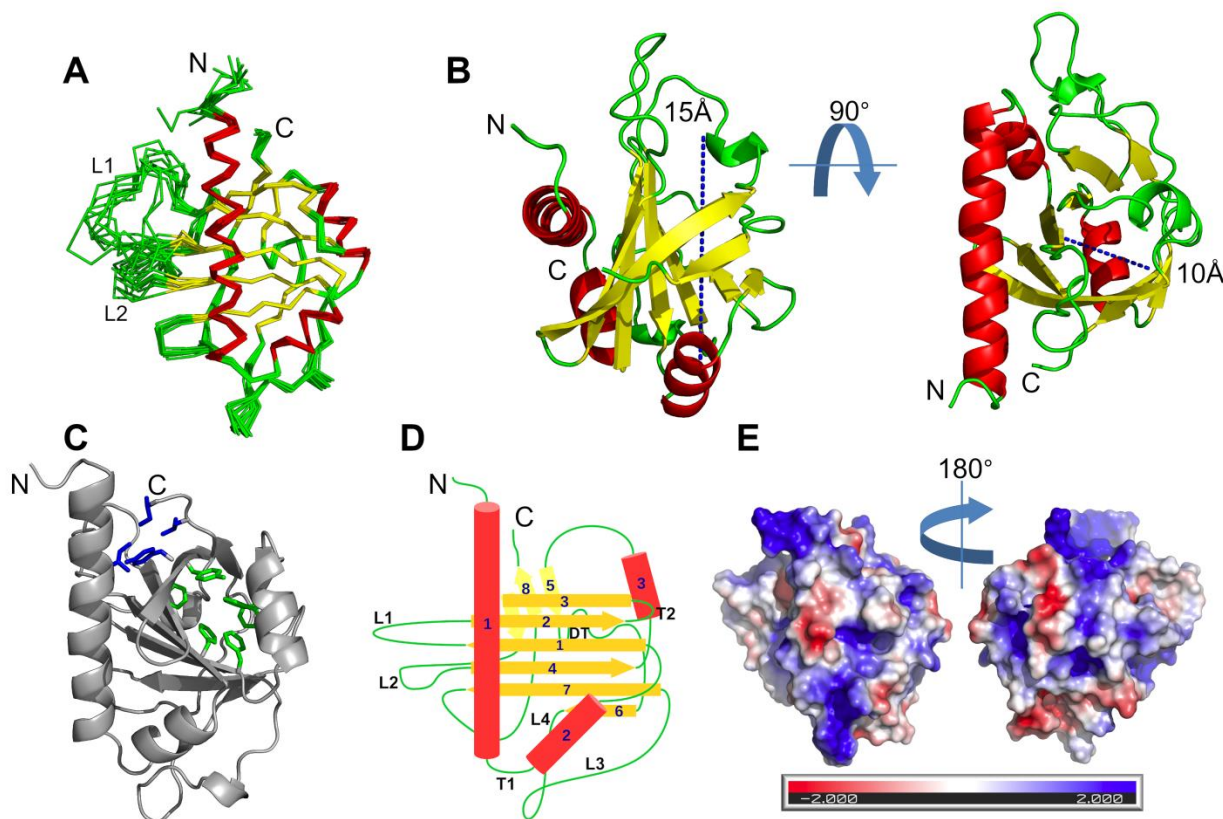


Figure 2: Three-dimensional structure of BPSL1445-SYLF (A) Ribbon ($C\alpha$) representation of BPSL1445-SYLF NMR solution structure shown as an ensemble of 10 conformers overlaid on the backbone atoms of ordered residues. The helices, the β -strands and the loops are in red yellow and green, respectively. (B) Cartoon representation of BPSL1445-SYLF in two different orientations; β -barrel length and diameter are indicated. (C) Cartoon representation of BPSL1445-SYLF with the side-chains of the aromatic core F75, Y104, F124, F126, F135, W141 (green) and L193, I195, I44, Y103 (blue). (D) Topology of BPSL1445-SYLF structure with the β -strands and α -helices in yellow and red, respectively. Loops (L), turns (T) and double turn (DT) are indicated. (E) Two orientations of BPSL1445-SYLF electrostatic potential surface.

The structural compactness, ascribable to the extensive network of hydrophobic interactions, is in line with the high melting temperature of BPSL1445-SYLF (Figure 1C). Most of the fold is well-defined in the bundle of structures, with a root-mean-square deviation (RMSD) of 0.58 ± 0.06 Å on the backbone atoms of residues N37-D77, N91-S108, K120-K192 (Table 1). Only the two protruding loops L1 and L2 are disordered (Figure 2A), as assessed by the paucity of NOEs and by the low heteronuclear NOE values in this region (see below). Finally, the electrostatic

surface of BPSL1445-SYLF (pI=9.6) does not present any remarkable feature, with negatively and positively charged patches evenly distributed on the protein surface (Figure 2E).

Table 1

Backbone Dynamics of BPSL1445-SYLF domain

We next assessed the dynamic properties of the backbone amides of BPSL1445-SYLF domain on the picosecond to nanosecond time scale based on ^{15}N relaxation data recorded on a 600 MHz spectrometer (Figure 3A-F).

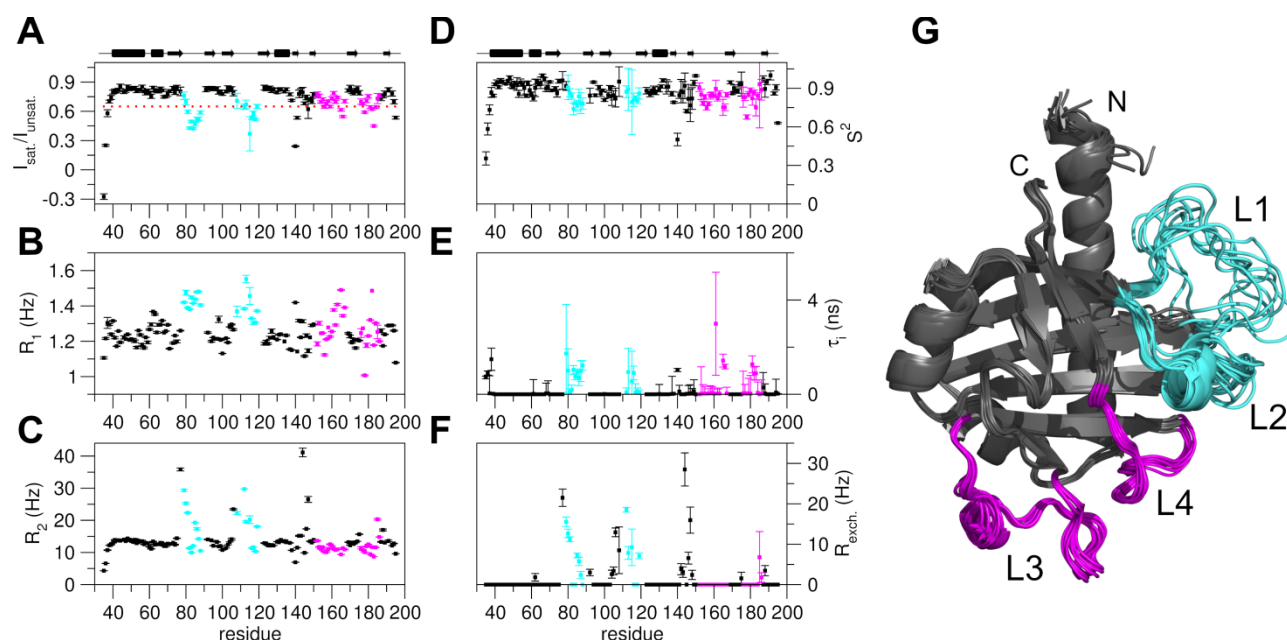


Figure 3: NMR dynamics of BPSL1445-SYLF. Plots of the experimental (A) heteronuclear NOE (the red line indicates the threshold value 0.65), (B) R_1 and (C) R_2 values of ^{15}N BPSL1445-SYLF backbone amides. Model-free analysis of the ^{15}N relaxation data with (D) order parameters S^2 , (E) internal correlation times τ_i , (F) exchange terms R_{ex} . Values associated to the loops are colored according to the coloring scheme represented in (G). (G) Superposition of the 10 lowest energy structures of BPSL1545-SYLF with the flexible (L1, L2 in cyan) and ordered loops (L3, L4 in magenta) highlighted.

High heteronuclear NOE values (above 0.7), indicative of restricted motion on the picosecond to nanosecond time scale were observed for residues adopting well-defined secondary structure conformations (Figure 3A). Also, residues within the well-defined loops L3 and L4 that form part of the β -barrel, displayed relatively high heteronuclear NOE values (between 0.6 and 0.7). Conversely, residues within L1 and L2 (Figure 3G), in agreement with the paucity of

intramolecular NOEs in this region, displayed a reduction of the heteronuclear NOE values, indicative of large-amplitude backbone motions on the sub-nanosecond time scale (Figure 3A). The increased R_2 values within L1, L2, with respect to the uniform R_1 and R_2 profiles along the secondary structure elements, suggested the presence of both fast internal motions (ps-ns) and chemical exchange in these regions (Figure 3B,C). Anomalous R_2 values in the C-terminus of $\beta 3$ and in the DT region, ascribable to slow (ms) chemical/conformational exchange, are likely due to unstable hydrogen bonds within the DT and $\beta 3$. R_2/R_1 values, properly filtered to exclude residues with high degree of internal mobility or ^{15}N line broadening²⁷, were used to estimate the overall rotational correlation time (τ_c) in model-free analysis. Its value ($\tau_c = 9.8 \pm 0.3$ ns) was well in line with the 17kDa molecular weight of the monomeric protein (Supplementary Figure S2C). Model-free analysis under conditions of anisotropic tumbling and based on the model selection criteria implemented in TENSOR2²⁸, yielded an order parameter (S^2) average value of 0.90 ± 0.03 (Figure 3D), whereby the patterns of the S^2 and exchange rate (R_{exch}) values mirrored the ^{15}N $\{^1\text{H}\}$ -NOE and R_2 values, respectively (Figure 3E,F).

Sequence and structural conservation in the SYLF family

The region of BPSL1445 downstream of the predicted signal peptide is annotated in public resources as being part of the SYLF domain family. Interestingly, the N-terminal helix and the first three β -strands observed in the BPSL1445-SYLF structure fall outside of the SYLF family boundaries as defined in the Pfam and InterPro databases. We thus suggest to adjust the N-terminal boundaries of such families in agreement with the newly determined SYLF domain structure. We also note that in several public databases the SYLF family features a C-terminal extension predicted to be mostly in an α -helical conformation, which is usually present in eukaryotes but absent in BPSL1445 (Supplementary Figure S1A). Using BPSL1445-SYLF sequence as a query we performed an iterative sequence search that identified more than 4,500 potential homologs in bacteria and eukaryotes. Most bacterial homologs, similarly to BPSL1445, contain a stand-alone SYLF domain. In Proteobacteria it is often preceded by an N-terminal signal peptide, which likely exports the protein to the periplasmic space or outer environment. Some eukaryotic members such as SH3YL1 and yeast Ysc84p/Lsb4p (LAS seventeen-binding protein 3) additionally contain an SH3 domain at the C-terminus, whereas family members from green plants and algae (stramenopiles) contain a FYVE zinc finger

domain at the N-terminus (Supplementary Figure S1A). Sequence conservation in the SYLF domain across different taxonomic groups comprises a group of glycines, mainly localized in loop regions and on $\beta 2$, few hydrophobic patches in the protein core (mostly located on $\beta 1$, $\beta 2$ and $\beta 4$) and few hydrophobic residues in the loops (Figure 4, Supplementary Figure S1B).

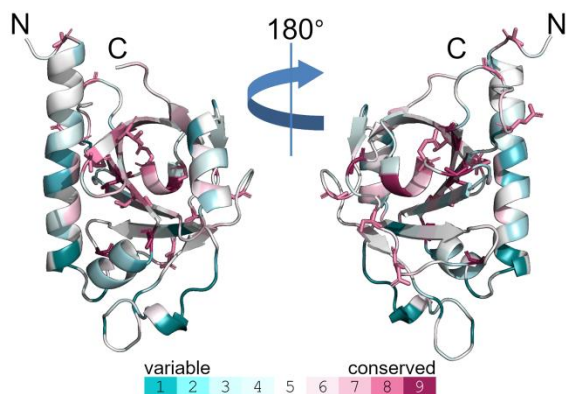


Figure 4: SYLF domain sequence conservation. ConSurf⁵³ analysis of BPSL1445-SYLF based on the alignment of 69 SYLF domain containing sequences from different taxonomic groups (Supplementary Figure S1B). Residues are colored by their conservation score using the grade color-coding bar. The sidechains of residues with conservation score ≥ 8 are shown.

Next, we used the BPSL1445-SYLF structure as query in a Dali structural similarity search³⁰ of the Protein Data Bank. The top match (Z-score=5.7) was Envelope integrity protein A (EipA)³¹ from *Brucella abortus*, a member of the Pfam family EipA (PF06577), whose structure (PDB entry: 5uc0) superimposes on the backbone atoms of BPSL1445-SYLF with a root mean square deviation (rmsd) of 3.7 Å (over 106 residues) with BPSL1445-SYLF (Supplementary Table S1). The two structures have similar topology and overlay relatively well on the β -barrel, with the most conserved residues localized on $\beta 1$, $\beta 2$, $\beta 4$, $\beta 7$ $\alpha 3$ and L1. The two major differences include: i. the length and the orientation of the N-terminal helix, that in EipA is shorter and tilted by 90° with respect to the corresponding helix in BPSL1445-SYLF (Figure 5A); notably within the SYLF family members this helix encompasses a region of high sequence variability (Figure 1A, Supplementary Figure S1A,B); ii. the double turn DT, which is replaced

in EipA by two β -hairpins (Figure 5B,C). Given that the SYLF and EipA domains share a similar fold, we further explored the possibility that they might be evolutionarily related. To this end, we performed HHpred³² profile-profile searches against the Pfam profile collection using three different multiple sequence alignments representing the SYLF family, that returned significant similarity scores with the EipA family, thus providing strong evidence of homology between the two domains (Supplementary Methods and Supplementary Figure S3). This is also in agreement with the SCOOP³³ predictions reported on the Pfam website. Collectively, we argue that sequence and structural similarities provide compelling evidence of a common evolutionary origin for the SYLF domain and EipA family. Functionally, EipA has been shown to determine cell envelope stress resistance and virulence in *Brucella abortus*³¹, while BPSL1445 does not seem to be involved in the virulence of *B. Pseudomalle*³⁴. Nonetheless, the establishment of a likely evolutionary link between the two domains might help leveraging any further information that might become available in the future regarding their cellular and molecular functions.

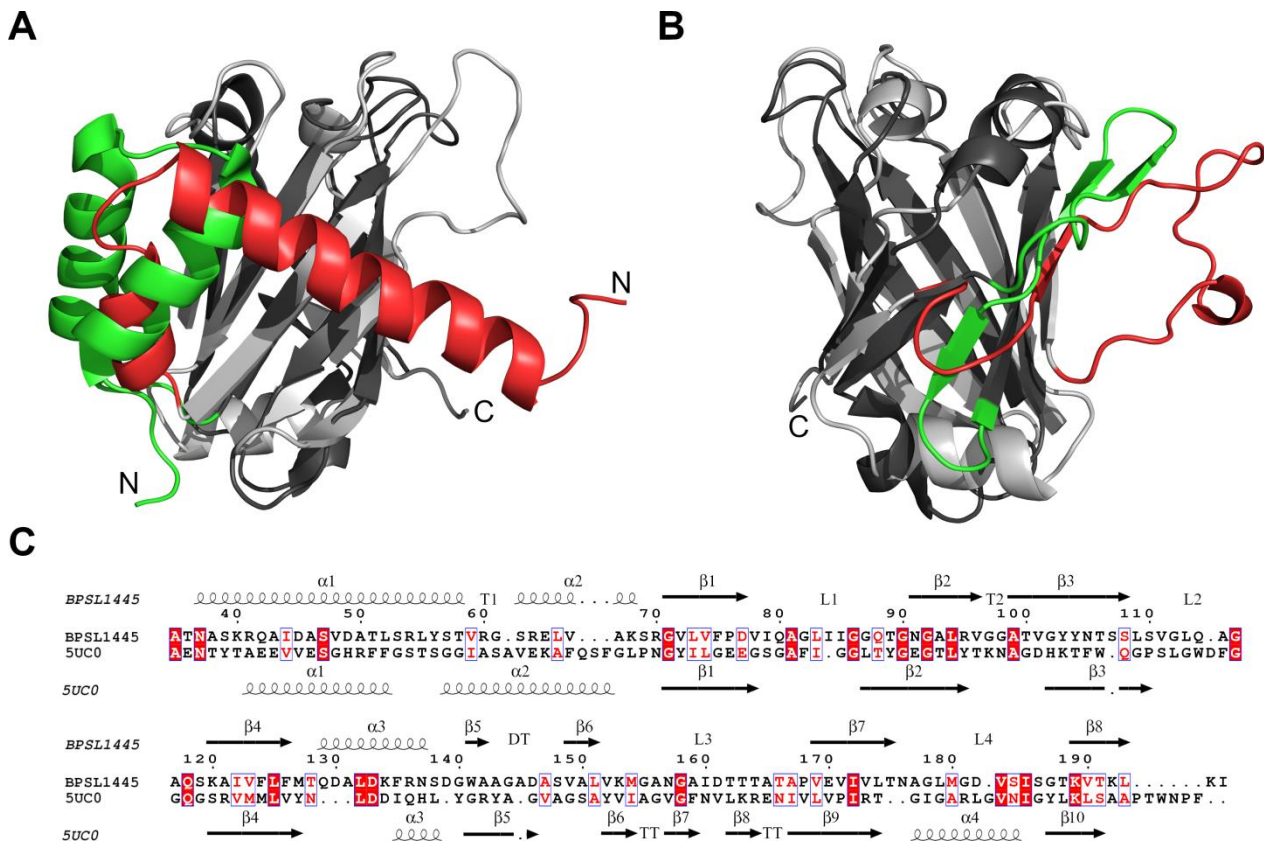


Figure 5: SYLF domain structural conservation. (A,B) Superposition of BPSL1445-SYLF (grey and red) and EipA (5uc0) structures (black and green) in two different orientations; for sake of clarity in (A) the regions A143-P168 (DT- $\beta 6$ -L3) of BPSL1445-SYLF and G146-I167 ($\beta 6$ -TT- $\beta 7$ - $\beta 8$ -TT) of EipA are not displayed; in (B) helices $\alpha 1$ and $\alpha 2$ of BPSL1445-SYLF and EipA are not displayed. (C) Structure guided sequence alignment (ESPrInt)⁵² as determined by DALI³⁰, elements of secondary structure of BPSL1445-SYLF and of EipA are represented at the top and at the bottom, respectively.

BPSL1445-SYLF binds to phosphatidylinositols *in vitro*

The SYLF domain of human SH3YL1 has been previously shown to interact with phosphoinositides³, we thus wondered whether also BPSL1445-SYLF domain might be a phospholipid-binding module. To investigate this possibility, we incubated purified GST-tagged BPSL1445-SYLF (Figure 6A) and GST alone as negative control (Supplementary Figure S4A) with phosphatidyl-inositol phosphate (PiP) strip membranes spotted with different lipids, and detected binding using anti-GST antibodies.

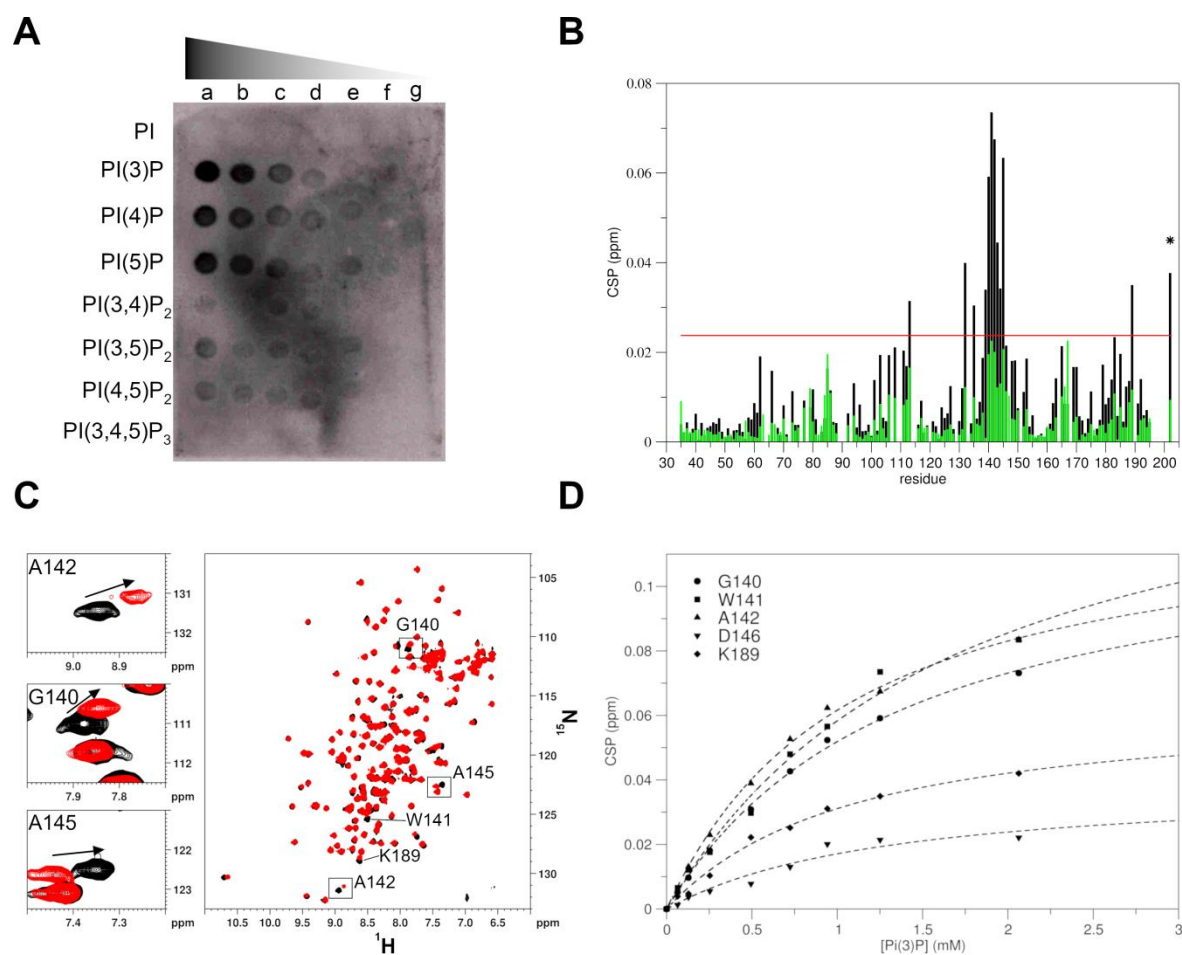


Figure 6: Binding of BPSL1445-SYLF to PiPs. (A) PiP-strip assay on BPSL1445-SYLF. The lipids concentration gradient spans from 100 pmol/spot to 1.6 pmol/spot. (B) Histograms showing residue-specific CSPs of ¹⁵N-labeled BPSL1445-SYLF (0.1 mM) upon addition of 10-fold excess of Pi(3)P (black) and Pi (green). Missing data correspond to prolines or overlapping resonances. The red line represents the average CSP value plus the standard deviation for Pi (3)P titration. The asterisk indicates the indolic NH of W141. (C) Overlay of the ¹H-¹⁵N HSQC spectrum of BPSL1445-SYLF without (black) and with 10-fold excess of Pi(3)P (red), on the left zoom view of selected spectral regions, with arrows indicating the direction of peak displacement upon binding. (D) Weighted average of G140, W141, A142, D146, K189, amides ¹H and ¹⁵N CSPs in the presence of increasing concentrations of Pi(3)P.

Indeed, GST-BPSL1445-SYLF was able to bind a range of phospholipids including phosphatidylinositol mono- and di-phosphates, with a slight preference for mono-phosphates. Conversely, little or no binding was observed to phosphatidylinositol-three-phosphates and phosphatidylinositol (Pi) (Figure 6A), respectively. BPSL1445-SYLF appears to specifically recognize PiPs, as no interaction was detected using membranes spotted with other important biological lipids (Supplementary Figure S4B,C). We further validated the binding of phosphoinositides by NMR chemical shift perturbation (CSP) experiments. Titration of ^{15}N -labeled BPSL1445-SYLF with water-soluble di-C8 forms of PiP monophosphates (Pi(3)P, Pi(4)P, Pi(5)P), di-phosphate (Pi(3,5)P₂) and three-phosphate (Pi(3,4,5)P₃) induced the perturbation of a discrete set of resonances, supporting a direct interaction between the lipid and the domain. Significant chemical shift changes independently from phosphorylation position were observed for (L132, F135, D139-A145, K189) (Figure 6B,C, Supplementary Figure S4B-F). In line with the results observed in PiP strip assays, the entity of the CSPs generated by phosphatidylinositol (Pi) was much lower as compared to the phosphorylated analogs, suggesting a reduced binding of Pi with respect to PiPs (Figure 6B). Indeed, non-linear fitting of the weighted average of ^1H and ^{15}N chemical-shift changes of selected BPSL1445-SYLF residues as a function of added lipids yielded dissociation constants (Kds) of 0.9 ± 0.3 mM for Pi(3)P (Figure 6D) and of 1.25 ± 0.4 mM and of 2.9 ± 1.0 mM for Pi(5)P and Pi(4)P, respectively. Binding of Pi(3,5)P₂ (7.0 ± 2.8 mM) and of Pi(3,4,5)P₃ (>10 mM) was weaker, confirming a preference for mono-phosphorylated PiP. Conversely, the shifts induced by Pi were too small to be fit with a binding isotherm, supporting the important role of the phosphate group to obtain productive binding. Mapping on the three-dimensional structure of BPSL1445-SYLF domain of the residues with the largest CSPs upon mono-phosphate phosphatidylinositols addition revealed that the effect of lipid binding is restricted to a well-defined area involving residues on $\alpha 3$ (L132, F135), $\beta 5$ and the double turn DT (D139-A145) and $\beta 8$ (K189) (Figures 1A, 7A). We next used the Schrödinger suite 2020-4 (Schrödinger, LLC, New York, NY, 2020) to generate a docking model of BPSL1445-SYLF in complex with di-C4 Pi(3)P, the latter displaying similar CSPs as di-C8 Pi(3)P (Supplementary Figure S4H). Among the best 50 docking poses, we selected the one with the best agreement with the CSPs data (Figure 7A, B), with the phosphatidyl group interacting with the side chains of K189 and of D139 (Figure 7C), and with one acyl chain exposed to the solvent and the other establishing weak

apolar interactions with the protein surface (Figure B,C, Supplementary Figure S5A). The binding pose was compatible with the electrostatic surface of the domain (Figure 7D). We also mutated into alanine four BPSL1445-SYLF charged residues within the predicted interaction surface (K134A, K189A, D139A, D146A) (Supplementary Figure S5A). Both K134A and K189A aggregated and could not be tested in binding experiments, whereas D139A and D146A were soluble and folded. Interestingly, di-C8-Pi(3)P titrations into the D146A mutant resulted in increased CSP (associated to a slightly better K_d ~0.7 mM) (Supplementary Figure S5B), conceivably due to favorable hydrophobic interactions between the alanine methyl group and the hydrophobic acyl chain. Alanine mutation of D139 (Supplementary Figure S5C) resulted only in a small reduction of CSPs, with a slight decrease in the affinity (K_d ~2 mM). Collectively, these results are compatible with the presented docking pose.

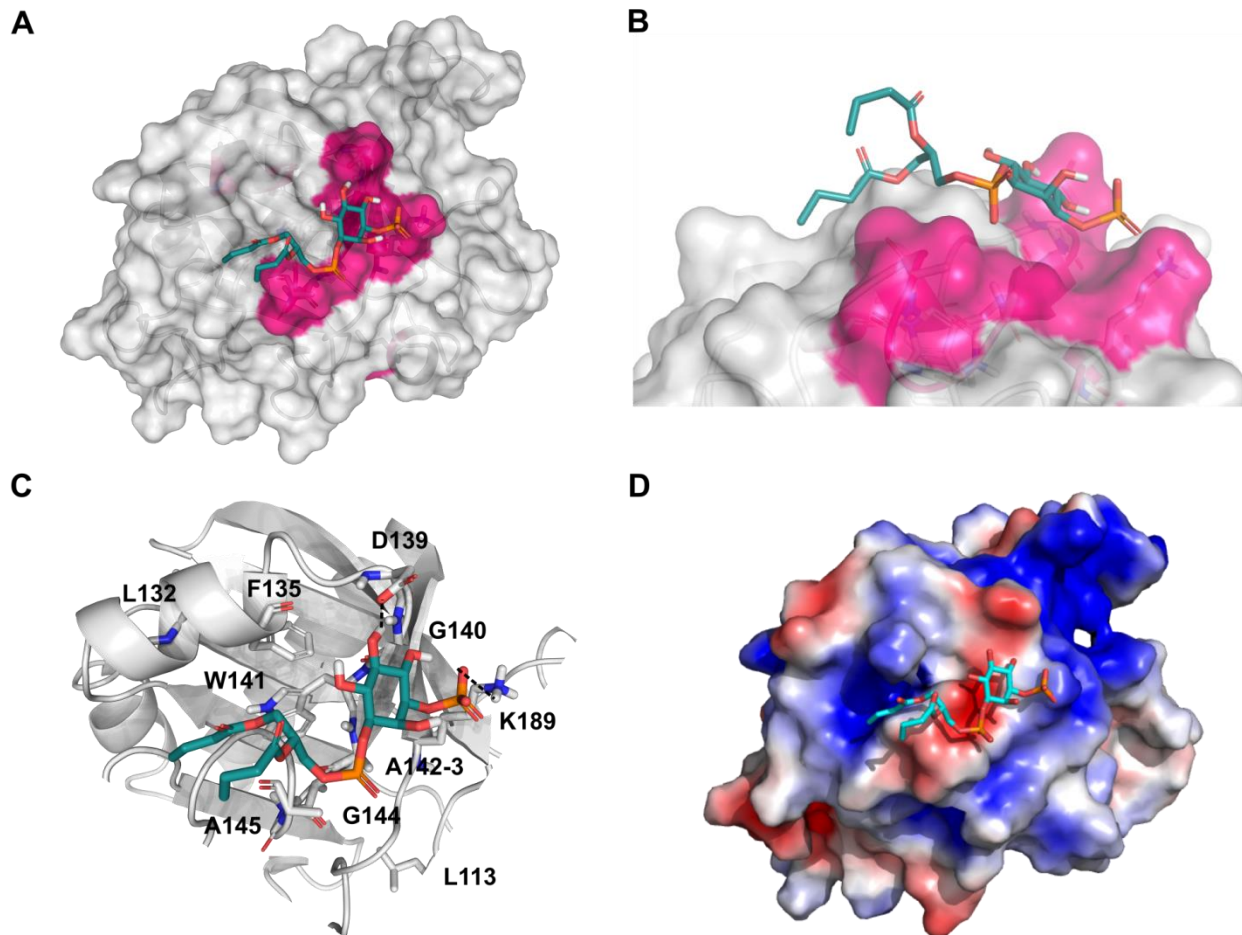


Figure 7: Model of BPSL1445-SYLF-Pi(3)P interaction. (A) Docking model of BPSL1445-SYLF (gray) in complex with di-C4-Pi(3)P (stick representation), residues with significant CSPs in magenta. (B) Same as (A) in a different orientation. (C) Zoom view of the binding site, with residues (sticks) with significant CSPs explicitly labeled. Hydrogen bonds are represented with dotted lines. (D) Electrostatic potential surface of the docking pose of BPSL1445-SYLF with di-C4-Pi(3)P (stick representation).

ral
ure

measured by CD spectroscopy (Figure 1C). Protein thermostability reduction upon PiP-binding is not uncommon and has been already observed in other PiP-binding domains, such as the VHS domain of Tom1 protein³⁵. Conceivably, binding of phosphatidylinositol induces conformational changes of BPSL1445-SYLF, that reduce structure compactness, eventually resulting in a lower T_m. The PiP binding potential of BPSL1445-SYLF is in line with the ability of secreted peripheral bacterial proteins to target the host cell membranes through PiP-mediated interactions^{6,36} exploiting binding motifs without significant sequence or structural similarity to the classical eukaryotic PiP-binding domains. Affinity of BPSL1445-SYLF to PiP is low and other factors, such as the basic residues located adjacent to the PiP binding site (K134, K192, K194) may contribute to membrane recruitment, as already observed for other lipid binding domains, like the PH domain^{37,38}.

The identified PiP-interaction site on BPSL1445-SYLF appears unique and could not have been predicted on the basis of sequence or structural similarities with other PiP-binding domains. However, similarly to the majority of the known PiP recognition domains, the interaction with the ligand does not involve a deep binding pocket, but is rather superficial. Interestingly, the region mostly affected by PiP binding in terms of CSPs is relatively well conserved in β -proteobacteria (Supplementary Figure S1C,D), suggesting a conserved function. Interestingly, this region does not coincide with the one suggested for human SH3YL1³. In the latter case the PiP binding site was mainly located on the N-terminal helix, that presents low sequence conservation between prokaryotes and eukaryotes (Figure 1A). However, in the same paper the authors pointed out that the whole domain had PiP binding activity, thus allowing for the possibility that an additional region of the human SH3YL1-SYLF domain might contribute to PiP interaction.

Finally, as the SYLF domain of the yeast protein Ysc84 has actin binding ability^{1,2}, we also explored the possibility that BPSL1445-SYLF could interact with F-actin. However, titration of F-actin into 0.13 mM ¹⁵N BPSL1445-SYLF did not induce appreciable perturbations in terms of CSPs and/or intensity changes (excluding the dilution effect) in the ¹⁵N HSQC spectrum of BPSL1445-SYLF, suggesting the absence of specific interactions under these conditions (Supplementary Figure S6).

Conclusions

In this study we provided first structural insights into the SYLF domain, an evolutionary conserved domain present both in prokaryotes and eukaryotes. We solved the solution structure and the dynamics of the SYLF domain of BPSL1445, a seroreactive antigen of *B. pseudomallei*, the pathogen responsible of the infectious disease melioidosis. The BPSL1445-SYLF domain folds into a very thermostable structure formed by a β -barrel core, with two flexible loops (L1 and L2) protruding out of the barrel and three α -helices packing against its surface. Similarly to what has been observed in reverse structural vaccinology studies from other seroreactive antigens of *B. pseudomallei*^{14,15,18}, these flexible loops might be profitably exploited as immunogenic epitopes, to elicit the production of antibodies for passive immunization. Structural comparison, supported by profile-profile alignments, indicates that SYLF is evolutionary related to the bacterial EipA domain. Previous studies attributed PiP-binding ability to the SYLF domain of human SH3YL1³, here we have shown that a prodrome of this function can be detected in prokaryotes, with very weak and promiscuous interactions between a *B. pseudomallei* SYLF domain and phosphatidylinositols. The binding involves a conserved domain surface in β -proteobacteria. We hypothesize that the PiP binding properties of the BPSL1445-SYLF domain, together with other bacterial effector molecules, might be exploited by the pathogen to target the membrane of the host cell. Further studies should be dedicated to the effective ability of BPSL1445 to interact with eukaryotic cell membranes and to characterize the PiP binding ability of the SYLF domain in higher organisms.

Methods

Protein expression and purification

Isotopically enriched BPSL1445 ($^{15}\text{N}/^{13}\text{C}$, or ^{15}N labelled) was expressed and purified as described previously³⁹. NMR samples were prepared in a buffer containing 150 mM NaCl, 20mM $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$, 2 mM DTT buffer at pH 6.3 with 0.16 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) and 10% D_2O .

NMR measurements

NMR experiments were performed at 295K on a 600-MHz spectrometer (Bruker Avance 600 Ultra Shield TM Plus, Bruker BioSpin) equipped with a triple-resonance TCI cryoprobe with a z-shielded pulsed-field gradient coil. Data were processed with Topspin3.6 (Bruker Biospin) and/or with NMRPipe⁴⁰ and analyzed with CCPN analysis 2.4 software⁴¹. The ^1H , ^{13}C and ^{15}N chemical shifts of backbone and side chains resonances have been assigned as described previously (BMRB:25387)³⁹. Proton–proton distance constraints were obtained from ^{15}N and ^{13}C separated 3D NOESY spectra employing 100 ms mixing times. $^3\text{J}(\text{HN}, \text{Ha})$ coupling constants were measured to derive restraints for ϕ dihedral angles. Additional ϕ/ψ restraints were obtained from backbone chemical shifts using Talos-N⁴². Hydrogen bond restraints were defined from slow-exchanging amide protons identified after $\text{H}_2\text{O}/\text{D}_2\text{O}$ exchange. ^1H – ^{15}N residual dipolar couplings were measured in isotropic and anisotropic phases obtained using polyacrylamide gel (10% w/v)⁴³. Heteronuclear $\{^1\text{H}\}^{15}\text{N}$ nuclear Overhauser enhancements (hetNOEs), longitudinal and transversal ^{15}N relaxation rates (R_1 , R_2) were measured using standard 2D methods⁴⁴, duty-cycle heating compensation were used for both T_1 and T_2 relaxation experiments⁴⁵. T_1 and T_2 decay curves were sampled at 12 different time points (50–2000 ms and 12–244 ms, respectively) collected in random order, with 2.5 seconds recovery delay. The $\{^1\text{H}\}^{15}\text{N}$ NOEs were measured recording HSQC⁴⁴ spectra with and without proton saturation in an interleaved fashion using a 4 seconds recycle delay. T_1 , T_2 and hetNOE values have been obtained using the NMRView fitting routine⁴⁶.

For NMR titrations with different phospholipids, at each titration point a 2D water-flip-back ^1H – ^{15}N -edited HSQC spectrum was acquired with 2048 (160) complex points for ^1H (^{15}N), respectively, apodised by 90° shifted squared (sine) window functions and zero filled to 256 points for indirect dimension. Titrations of 10 mM stock solutions (in NMR buffer) of Pi and PiPs

(Pi3P, Pi4P, Pi5P, Pi(3,5)P₂, Pi(3,4,5)P₃ (Echelon Biosciences Inc., Salt Lake City, Utah) were performed adding 0.5, 1, 2, 4, 6, 8, 11, 20 equivalents of ligands into the ¹⁵N BPSL1445 (0.13mM; in NMR buffer). Assignment of the labeled proteins in the presence of the ligands was obtained following individual cross-peaks through the titration series. For each residue the weighted average of the ¹H and ¹⁵N chemical shift perturbation (CSP) was calculated as $CSP = [(\Delta H N^2 + \Delta N^2/25)/2]^{1/2}$ ⁴⁷. The apparent dissociation constants of PiPs-¹⁵N-BPSL1445, interactions were estimated from least-squares fitting of CSPs as a function of total ligand concentration according to the equation:

$$\delta_i = \frac{b - \sqrt{b^2 - 4ac}}{2a}$$

with $a = (K_a/\delta_b)[P_t]$, $b = 1 + K_a([L_{ti}] + [P_t])$, and $c = \delta_b K_a [L_{ti}]$, where δ_i is the absolute change in chemical shift for each titration point, $[L_{ti}]$ is the total ligand concentration at each titration point, $[P_t]$ is the total protein concentration, $K_a = 1/K_d$ is the association constant, and δ_b is the chemical shift of the resonance in the complex. The final K_d s and the corresponding errors were the average plus one standard deviation of the fitting of those residues with significant CSPs (>average plus one standard deviation). Details on NMR-titrations with F-actin are reported in Supplementary information.

Structure calculation

Structures were calculated using ARIA 2.3.2⁴⁸ combined to CNS⁴⁹ using the experimentally derived restraints (Table 1). All NOEs were assigned manually and calibrated by ARIA 2.3.2. A total of eight iterations (20 structures in the first six iterations) were performed: 200 structures were computed in the last iteration. The ARIA 2.3.1 default water refinement was performed on the 40 best structures of the final iteration. Structural quality was assessed using PSVS server⁵⁰ and PROCHECK-NMR⁵¹. The family of the 10 lowest energy structures has been deposited in the Protein Data Bank with the accession code 7OFN.

Circular dichroism measurements

CD spectra (20°C) were acquired on a Jasco J-815 using a rectangular quartz cuvette (1 mm path length, Hellma). Each spectrum was averaged over 4 scans collected in 0.1 nm intervals with an average time of 0.5 s. BPSL1445 concentration was 10 μM in 150 mM NaF, 20 mM

NaH₂PO₄/Na₂HPO₄ buffer (pH 7). Thermal denaturation curves (in the absence and presence of 20 fold excess of Pi3P) were obtained monitoring the ellipticity at 220 nm, 215 nm and at 209 nm from 20 °C to 95 °C at a constant scan rate of 1 C°/min.

PIP strip binding assays

Binding of GST- BPSL1445-SYLF and of GST alone to Membrane PiP strips and Lipid-strips (Echelon Biosciences, Inc., Salt Lake City, Utah) was performed according to the manufacturer instructions. Briefly, membrane PiP strips were blocked overnight at 4°C with 3% of BSA (Sigma) dissolved in PBS buffer (Euroclone) containing 0.1% Tween-20 (TPBS) and then incubated for 1 h at room temperature (RT) with the recombinant proteins (5µM). Membranes were washed 3 times with TPBS and incubated for 1 h at room temperature with anti-GST antibody (1:2000) (Merck). After 3 washes with TPBS, membranes were incubated for 1 h with a peroxidase-conjugated affinipure goat anti-rabbit IgG (1:5000; Biorad). After 3 washes with TPBS, bound proteins were detected by ECL (GE Healthcare).

Sequence Searches, multiple sequence alignment and Docking

Details on sequence searches, multiple sequence alignments and docking calculations are described in Supplementary Information. Molecular images and electrostatic potential were generated by the open source PyMOL Molecular Graphics System, Version 2.4 Schrödinger, LLC.

Acknowledgements

This project was supported by Fondazione CARIPLO (Progetto Vaccini, contract number 2009-3577) and Associazione Italiana Ricerca sul Cancro (AIRC, IG-214400)

Supporting Information: Supplementary Methods and Figures. Multiple sequence alignment files.

Bibliography

- (1) Robertson, A. S., Allwood, E. G., Smith, A. P. C., Gardiner, F. C., Costa, R., Winder, S. J., and Ayscough, K. R. (2009) The WASP Homologue Las17 Activates the Novel Actin-regulatory Activity of Ysc84 to Promote Endocytosis in Yeast. *Mol. Biol. Cell* (Chang, F., Ed.) *20*, 1618–1628.
- (2) Urbanek, A. N., Allwood, E. G., Smith, A. P., Booth, W. I., and Ayscough, K. R. (2015) Distinct Actin and Lipid Binding Sites in Ysc84 Are Required during Early Stages of Yeast Endocytosis. *PLoS One* (Arkowitz, R. A., Ed.) *10*, e0136732.
- (3) Hasegawa, J., Tokuda, E., Tenno, T., Tsujita, K., Sawai, H., Hiroaki, H., Takenawa, T., and Itoh, T. (2011) SH3YL1 regulates dorsal ruffle formation by a novel phosphoinositide-binding domain. *J. Cell Biol.* *193*, 901–916.
- (4) Sutipatanasomboon, A., Herberth, S., Alwood, E. G., Häweker, H., Müller, B., Shahriari, M., Zienert, A. Y., Marin, B., Robatzek, S., Praefcke, G. J. K., Ayscough, K. R., Hülskamp, M., and Schellmann, S. (2017) Disruption of the plant-specific CFS1 gene impairs autophagosome turnover and triggers EDS1-dependent cell death. *Sci. Rep.* *7*, 8677.
- (5) Balla, T. (2005) Inositol-lipid binding motifs: signal integrators through protein-lipid and protein-protein interactions. *J. Cell Sci.* *118*, 2093–2104.
- (6) Pemberton, J. G., and Balla, T. (2019) Polyphosphoinositide-Binding Domains: Insights from Peripheral Membrane and Lipid-Transfer Proteins, in *Advances in Experimental Medicine and Biology*, pp 77–137.
- (7) Kutateladze, T. G. (2010) Translation of the phosphoinositide code by PI effectors. *Nat. Chem. Biol.* *6*, 507–513.
- (8) Currie, B. J., Dance, D. A. B., and Cheng, A. C. (2008) The global distribution of *Burkholderia pseudomallei* and melioidosis : an update. *Trans. R. Soc. Trop. Med. Hyg.* *102*, S1–S4.
- (9) Wiersinga, W. J., Virk, H. S., Torres, A. G., Currie, B. J., Peacock, S. J., Dance, D. A. B., and Limmathurotsakul, D. (2018) Melioidosis. *Nat. Rev. Dis. Prim.* *4*, 17107.
- (10) Wiersinga, W. J., Poll, T. Van Der, White, N. J., Day, N. P., and Peacock, S. J. (2006) Melioidosis : insights into the pathogenicity of *Burkholderia pseudomallei* *4*, 272–282.
- (11) Cheng, A. C., and Currie, B. J. (2005) Melioidosis : Epidemiology , Pathophysiology , and Management *18*, 383–416.
- (12) Felgner, P. L., Kayala, M. A., Vigil, A., Burk, C., Nakajima-sasaki, R., Pablo, J., Molina, D. M., Hirst, S., Chew, J. S. W., Wang, D., Tan, G., Duffield, M., Yang, R., Neel, J., Chantratita, N., Bancroft, G., Lertmemongkolchai, G., Davies, D. H., and Baldi, P. (2009) A *Burkholderia pseudomallei* protein microarray reveals serodiagnostic and cross-reactive antigens. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 13499–13504.
- (13) Suwannasaen, D., Mahawantung, J., Chaowagul, W., Limmathurotsakul, D., Felgner, P. L., Davies, H., Bancroft, G. J., Titball, R. W., and Lertmemongkolchai, G. (2011) Human Immune Responses to *Burkholderia pseudomallei* Characterized by Protein Microarray

Analysis. *J. Infect. Dis.* 203, 1002–1011.

(14) Lassaux, P., Peri, C., Ferrer-Navarro, M., Gourlay, L. J., Gori, A., Conchillo-Solé, O., Rinchai, D., Lertmemongkolchai, G., Longhi, R., Daura, X., Colombo, G., and Bolognesi, M. (2013) A Structure-Based Strategy for Epitope Discovery in *Burkholderia pseudomallei* OppA Antigen. *Structure* 21, 167–175.

(15) Gourlay, L. J., Peri, C., Ferrer-Navarro, M., Conchillo-Solé, O., Gori, A., Rinchai, D., Thomas, R. J., Champion, O. L., Michell, S. L., Kewcharoenwong, C., Nithichanon, A., Lassaux, P., Perletti, L., Longhi, R., Lertmemongkolchai, G., Titball, R. W., Daura, X., Colombo, G., and Bolognesi, M. (2013) Exploiting the *Burkholderia pseudomallei* acute phase antigen BPSL2765 for structure-based epitope discovery/design in structural vaccinology. *Chem. Biol.* 20, 1147–1156.

(16) Gourlay, L. J., Thomas, R. J., Peri, C., Conchillo-Solé, O., Ferrer-Navarro, M., Nithichanon, A., Vila, J., Daura, X., Lertmemongkolchai, G., Titball, R., Colombo, G., and Bolognesi, M. (2015) From crystal structure to in silico epitope discovery in the *Burkholderia pseudomallei* flagellar hook-associated protein FlgK. *FEBS J.* 282, 1319–33.

(17) Nithichanon, A., Rinchai, D., Gori, A., Lassaux, P., Peri, C., Conchillo-Solé, O., Ferrer-Navarro, M., Gourlay, L. J., Nardini, M., Vila, J., Daura, X., Colombo, G., Bolognesi, M., and Lertmemongkolchai, G. (2015) Sequence- and Structure-Based Immunoreactive Epitope Discovery for *Burkholderia pseudomallei* Flagellin. *PLoS Negl. Trop. Dis.* 9, e0003917.

(18) Gaudesi, D., Peri, C., Quilici, G., Gori, A., Ferrer-Navarro, M., Conchillo-Solé, O., Thomas, R., Nithichanon, A., Lertmemongkolchai, G., Titball, R., Daura, X., Colombo, G., and Musco, G. (2015) Structure-Based Design of a B Cell Antigen from *B. pseudomallei*. *ACS Chem. Biol.* 10, 803–812.

(19) Rodrigues, F., Sarkar-Tyson, M., Harding, S. V., Sim, S. H., Chua, H. H., Lin, C. H., Han, X., Karuturi, R. K. M., Sung, K., Yu, K., Chen, W., Atkins, T. P., Titball, R. W., and Tan, P. (2006) Global Map of Growth-Regulated Gene Expression in *Burkholderia pseudomallei*, the Causative Agent of Melioidosis. *J. Bacteriol.* 188, 8178–8188.

(20) Ooi, W. F., Ong, C., Nandi, T., Kreisberg, J. F., Chua, H. H., Sun, G., Chen, Y., Mueller, C., Conejero, L., Eshaghi, M., Ang, R. M. L., Liu, J., Sobral, B. W., Korbsrisate, S., Gan, Y. H., Titball, R. W., Bancroft, G. J., Valade, E., and Tan, P. (2013) The Condition-Dependent Transcriptional Landscape of *Burkholderia pseudomallei*. *PLoS Genet.* (Garsin, D. A., Ed.) 9, e1003795.

(21) Tuanyok, A., Tom, M., Dunbar, J., and Woods, D. E. (2006) Genome-Wide Expression Analysis of *Burkholderia pseudomallei* Infection in a Hamster Model of Acute Melioidosis. *Infect. Immun.* 74, 5465–5476.

(22) Babu, M. M., Priya, M. L., Selvan, A. T., Madera, M., Gough, J., Aravind, L., and Sankaran, K. (2006) A Database of Bacterial Lipoproteins (DOLOP) with Functional Assignments to Predicted Lipoproteins. *J. Bacteriol.* 188, 2761–2773.

(23) Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., and Marchler-Bauer, A. (2020) CDD/SPARCLE: The

conserved domain database in 2020. *Nucleic Acids Res.* 48.

(24) Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., and Bateman, A. (2021) Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49.

(25) Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., Brown, S. D., Chang, H. Y., El-Gebali, S., Fraser, M. I., Gough, J., Haft, D. R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., Natale, D. A., Necci, M., Nuka, G., Orengo, C., Pandurangan, A. P., Paysan-Lafosse, T., Pesseat, S., Potter, S. C., Qureshi, M. A., Rawlings, N. D., Redaschi, N., Richardson, L. J., Rivoire, C., Salazar, G. A., Sangrador-Vegas, A., Sigrist, C. J. A., Sillitoe, I., Sutton, G. G., Thanki, N., Thomas, P. D., Tosatto, S. C. E., Yong, S. Y., and Finn, R. D. (2019) InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* 47.

(26) Quilici, G., Berardi, A., Gaudesi, D., Gourlay, L. J., Bolognesi, M., and Musco, G. (2015) Backbone and side-chain ¹H, ¹⁵N, ¹³C assignment and secondary structure of BPSL1445 from *Burkholderia pseudomallei*. *Biomol. NMR Assign.* 9, 347–350.

(27) Barbato, G., Ikura, M., Kay, L. E., Bax, A., and Pastor, R. W. (1992) Backbone Dynamics of Calmodulin Studied by ¹⁵N Relaxation Using Inverse Detected Two-Dimensional NMR Spectroscopy: The Central Helix Is Flexible. *Biochemistry* 31, 5269–5278.

(28) Dosset, P., Hus, J. C., Blackledge, M., and Marion, D. (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR* 16, 23–28.

(29) Lakshmi, B., Sinduja, C., Archunan, G., and Srinivasan, N. (2014) Ramachandran analysis of conserved glycol residues in homologous proteins of known structure. *Protein Sci.* 23, 843–850.

(30) Holm, L., and Laakso, L. M. (2016) Dali server update. *Nucleic Acids Res.* 44, W351–W355.

(31) Herrou, J., Willett, J. W., Fiebig, A., Varesio, L. M., Czyż, D. M., Cheng, J. X., Ultee, E., Briegel, A., Bigelow, L., Babnigg, G., Kim, Y., and Crosson, S. (2018) Periplasmic protein EipA determines envelope stress resistance and virulence in *Brucella abortus*. *Mol. Microbiol.* mmi.14178.

(32) Gabler, F., Nam, S., Till, S., Mirdita, M., Steinegger, M., Söding, J., Lupas, A. N., and Alva, V. (2020) Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Curr. Protoc. Bioinforma.* 72.

(33) Bateman, A., and Finn, R. D. (2007) SCOOP: a simple method for identification of novel protein superfamily relationships. *Bioinformatics* 23, 809–814.

(34) Moule, M. G., Hemsley, C. M., Seet, Q., Guerra-Assunção, J. A., Lim, J., Sarkar-Tyson, M., Clark, T. G., Tan, P. B. O., Titball, R. W., Cuccui, J., and Wren, B. W. (2014) Genome-Wide Saturation Mutagenesis of *Burkholderia pseudomallei* K96243 Predicts Essential Genes and Novel Targets for Antimicrobial Development. *MBio* (Parkhill, J., Ed.) 5.

(35) Xiong, W., Tang, T.-X., Littleton, E., Karcini, A., Lazar, I. M., and Capelluto, D. G. S.

- (2019) Preferential phosphatidylinositol 5-phosphate binding contributes to a destabilization of the VHS domain structure of Tom1. *Sci. Rep.* **9**, 10868.
- (36) Ham, H., Sreelatha, A., and Orth, K. (2011) Manipulation of host membranes by bacterial effectors. *Nat. Rev. Microbiol.* **9**, 635–646.
- (37) Cohen, L. A., Honda, A., Varnai, P., Brown, F. D., Balla, T., and Donaldson, J. G. (2007) Active Arf6 Recruits ARNO/Cytohesin GEFs to the PM by Binding Their PH Domains. *Mol. Biol. Cell* (Schwartz, M. A., Ed.) **18**, 2244–2253.
- (38) Hammond, G. R. V., and Balla, T. (2015) Polyphosphoinositide binding domains: Key to inositol lipid biology. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **1851**, 746–758.
- (39) Quilici, G., Berardi, A., Gaudesi, D., Gourlay, L. J., Bolognesi, M., and Musco, G. (2015) Backbone and side-chain ¹H, ¹⁵N, ¹³C assignment and secondary structure of BPSL1445 from *Burkholderia pseudomallei*. *Biomol. NMR Assign.* **9**, 347–350.
- (40) Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**.
- (41) Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct. Funct. Genet.* **59**, 687–696.
- (42) Shen, Y., and Bax, A. (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–241.
- (43) Sass, H.-J., Musco, G., J. Stahl, S., T. Wingfield, P., and Grzesiek, S. (2000) Solution NMR of proteins within polyacrylamide gels: Diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *J. Biomol. NMR* **18**, 303–309.
- (44) Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (2005) Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by ¹⁵N NMR Relaxation. *Biochemistry* **33**, 5984–6003.
- (45) Yip, G. N. B., and Zuiderweg, E. R. P. (2005) Improvement of duty-cycle heating compensation in NMR spin relaxation experiments. *J. Magn. Reson.* **176**, 171–178.
- (46) Johnson, B. A., and Blevins, R. A. (1994) NMR View: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–14.
- (47) Grzesiek, S., Stahl, S. J., Wingfield, P. T., and Bax, A. (1996) The CD4 determinant for downregulation by HIV-1 Nef directly binds to Nef. Mapping of the Nef binding surface by NMR. *Biochemistry* **35**, 10256–10261.
- (48) Rieping, W., Bardiaux, B., Bernard, A., Malliavin, T. E., and Nilges, M. (2007) ARIA2: Automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* **23**, 381–382.

- (49) Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54, 905–921.
- (50) Bhattacharya, A., Tejero, R., and Montelione, G. T. (2006) Evaluating protein structures determined by structural genomics consortia. *Proteins Struct. Funct. Bioinforma.* 66, 778–795.
- (51) Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8, 477–86.
- (52) Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, W320–W324.
- (53) Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 44, W344–W350.

Table 1: Summary of conformational constraints and statistics for the 10 lowest energy structures of BPSL1445	
Restraints information ^{a,b}	<SA> ^c
Total number of experimental distance restraints	2809
NOEs (intraresidual/sequential/medium/long)	1133/554/359/763
Hydrogen bonds	59
Dihedral angle restraints (Φ/Ψ)	124/124
Residual dipolar couplings ^d	58
Deviation from idealized covalent geometry	
Bonds (Å)	0.0031 ± 0.0001
Angles (°)	0.413 ± 0.008
Coordinate r.m.s.deviation (Å)^e	
Ordered backbone atoms (N, C α , C')	0.580 ± 0.061
Ordered heavy atoms	1.061 ± 0.064
Ramachandran quality parameters (%)^f	
Residues in most favored regions	96.0%
Residues in allowed regions	3.8%
Global quality scores (Raw/Z)⁵⁰	
Verify3D	0.25/-3.37
ProsaII	0.50/ -0.62
Procheck	-0.02/0.24
MolProbity	10.32/-0.25

^a No distance restraint in any of the structures included in the ensemble was violated by more than 0.5 Å.

^b No dihedral angle restraints in any of the structures included in the ensemble was violated by more than 5°.

^c Simulated annealing, statistics refers to the ensemble of 10 structures with the lowest energy.

^d RDC restraints have been used only for residues adopting regular secondary structure.

^e Root mean square deviation from the lowest energy structure calculated on residues N37-D77, N91-S108, K120-K192.

^f Ramachandran quality parameters calculated on residues N37-D77, N91-S108, K120-K192