The Phage-shock-protein (PSP) Envelope Stress Response: Discovery of Novel Partners and Evolutionary History

Janani Ravi¹,²*, Vivek Anantharaman³, Samuel Zorn Chen⁴, Evan Pierce Brenner¹, Pratik Datta², L Aravind⁵*, Maria Laura Gennaro²*.
¹Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA; ²Public Health Research Institute, Rutgers University, Newark, NJ, USA; ³National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD, USA; ⁴Computer Science Engineering Program, Michigan State University, East Lansing, MI, USA. *Corresponding authors: janani.ravi@cuanschutz.edu; aravind@nih.gov; marila.gennaro@rutgers.edu

Table of Contents

Supplementary Figures and Tables

Supplementary Figures

Figure S1. Mutual exclusion phyletic pattern of PspM, Thioredoxin
Figure S2. PspA maximum likelihood gene tree with bootstrap values by RAxML-NG.

Supplementary Tables

Table S1. Summary of PSP partner domains (query proteins)
Table S2. Summary of novel PSP partner domains
Table S3. Domain architectures, genomic contexts, and lineages of representative PspA/Snf7 homologs. (next pages)
Table S4. Domain architectures, genomic contexts, and lineages of representative homologs of PSP cognate partner domains. (next pages)

Supplementary Text

1. Domain definitions
   1.1. LiaI-LiaF-TM and Toastrack domains
   1.2 PspM and PspN domains
   1.3 PspAA and PspAB domains

2. Novel PSP associations
   2.1 PspA/Snf7 domain architectures
   2.2 Vps4 and AAA⁺-ATPases
   2.3 PspA with PspM or Thioredoxin
   2.4 Novel contexts containing Toastrack
      Toastrack and TM domains
      Toastrack and transcription factors

References
Supplementary Figures and Tables
All PSP results (data summarizations and visualizations) can be accessed via our easy-to-use interactive web app: https://jravilab.org/psp.

Supplementary Figures

**Figure S1. Mutual exclusion phyletic pattern of PspM, Thioredoxin**
Figure S2. PspA maximum likelihood gene tree with bootstrap values by RAxML-NG.
## Supplementary Tables

### Table S1. Summary of PSP partner domains (query proteins)

| Domain Name | Old Name/Aliases | Pfam | PDB  |
|-------------|------------------|------|------|
| PspA        | PspA.IM30, LiaH, Vipp1, YjfJ, YdjF | PF04012 | 4WHE |
| Snf7        |                  | PF003357 | 5FD7 |
| PspA/Snf7   | PspA/ESCRT-III   | CL0235 | 3FRV |
| Toastrack   | DUF4097/DUF2154/DUF2807 LiaF/LiaG, YvlB, YthC, Toastrack_N (N-terminal region of Toastrack; PF17115) | PF13349/ PF09922/ PF10988/PF17115 | 4QRK/ 4OPW |
| Lial–LiaF-TM | DUF2157 Lia/LiaF | PF09925 | |
| PspM        | Rv2743c          |      | |
| PspN        | Rv2742c          |      | |
| PspN_N      | N-terminal region of PspN |      | |
| DUF3046     | C-terminal region of PspN | PF11248 | |
| PspB        |                  | PF06667 | |
| PspC        | YthA/YthB, YvlC  | PF04024 | |

### Table S2. Summary of novel PSP partner domains

| Domain Name | Old Name/Aliases | Pfam | PDB  |
|-------------|------------------|------|------|
| HAAS        | DUF1700 alpha-helical/ DUF1129/Yip1/ DUF1048 | PF08006/PF06570/ PF04893/PF06304 | 2O3L |
| SHOCT-bihelical | DUF1707, SHOCT | PF08044/PF09851 | |
| PspAA       | PspA-associated domain A |      | |
| PspAB       | PspA-associated domain B |      | |
| Vps4-AAA-ATPase/ Classical-AAA-ATPase | PF08432/ PF00004 | 5FVK/ 3U5Z |
| MIT         |                  | PF04212 | 5FVK |
| Domain Architecture                          | Genomic Context   | PF Number | Lineage |
|----------------------------------------------|-------------------|-----------|---------|
| Thioredoxin                                  |                   | PF00085   | 2OE3    |
| ClgR-HTH                                     | XRE-HTH/HTH\_3,  | PF01381   | 6IRP    |
|                                              | ClgR, cHTH        |           |         |
| TM-Flotillin dyad                            | Flot              | PF15975   |         |
| Band-7                                      | Band\_7           | PF01145   | 3BK6    |
| Spermine synthase                           | Spermine\_synth   | PF01564   | 6O63    |
| YjfL-TM(\_s)                                | DUF350            | PF03994   |         |
| CesT\_Tir                                   | CesT              | PF05932   | 1TTW    |
| CesT\_Tir-DUF2170                           | DUF2170           | PF09938   |         |
| TPM\_phosphatase                            |                   | PF04536   | 4OA3    |
| SHOCT\_like                                  | DUF1707           | PF08044   |         |
| Caspase                                     | Peptidase\_C14   | PF00656   | 3UO8    |
| PadR\_like\_wHTH                            | PadR-HTH/PadR     | PF03551   | 1XMA    |
| PadR\_wHTH                                  |                   |           |         |
| RHH                                         | TraJ-RHH/RHH\_1    | PF01402   | 3OD2    |
| SIGMA\_HTH                                   |                   |           |         |
| GerE\_HTH/DUF2089-HTH                       | GerE/ DUF2089     | PF00196/PF09862 | 2JPC |
| GNTR\_HTH                                   | GntR              | PF00392   | 4R1H    |

Table S3. Domain architectures, genomic contexts, and lineages of representative PspA/Snf7 homologs. (next pages)

Table S4. Domain architectures, genomic contexts, and lineages of representative homologs of PSP cognate partner domains. (next pages)
Supplementary Text

1. Domain definitions

1.1. LiaI-LiaF-TM and Toastrack domains
To best characterize the LiaIGF proteins, we used PSI-BLAST searches from the three sub-sequences of the full-length proteins, N-terminal TM region of LiaF, the C-terminus globular domain (DUF2154) of LiaF, and the globular domain in LiaG (DUF4097), followed by structure-informed sequence alignment.
These analyses revealed that LiaI and LiaG bear remarkable similarities to the N-terminal TM and C-terminal globular regions of the LiaF protein, respectively. We discovered that the globular domains of these LiaG-LiaF proteins are homologs of each other and that the profiles detected by Pfam in this region, DUF2154, DUF4097, and DUF2807, can be unified into a single domain called “Toastrack” that has a single-stranded right-handed beta-helix fold with a unique N-terminal 7-stranded region that displays a complex intertwining of strands (PDB: 4QRK; Toastrack_N, Pfam: PF17115 only has this unique N-terminal region) [Fig. 1A; Table S1]. Likewise, the homology between the 4TM (four TM) regions of LiaI and the N-terminal domain DUF2157 of LiaF led us to rename the 4TM region as ‘LiaI-LiaF-TM’ [Fig. 1B; Table S1]. Thus, the results of our analyses define two new domains: LiaI-LiaF-TM and Toastrack.

1.2 PspM and PspN domains
PspM comprises two TM regions and no other distinct domains. PspN contains a short domain at the C-terminus, DUF3046, and a yet uncharacterized N-terminal domain, which we now call PspN_N. We found that DUF3046 is α-helical with highly conserved threonine and cysteine residues that might be required for its function.
To further characterize the DUF3046 homologs, we used nucleotide sequences rather than translated open reading frames (ORFs), followed by sequence alignment analysis [Fig. 1B]. We found that the DUF3046 domain, which is widespread across Actinobacteria, is more similar to the short downstream protein, Rv2738c, than to the C-terminus of the fourth member of the M. tuberculosis operon, PspN (encoded by Rv2742c).

1.3 PspAA and PspAB domains
The PspA neighborhood analysis identified a new component in the proximity of PspA, which is a protein containing a novel trihelical domain (with absolutely conserved R and D) present in Euryarchaeota, Thaumarchaeota, Actinobacteria, Chloroflexi, Firmicutes, and a few Alpha- and Gammaproteobacteria). This protein occurs in a two-gene cluster with PspA [Fig. 3; Table S3; e.g., MA_1460; AAM04874.1, Methanosarcina]. This domain mostly occurs by itself but is occasionally fused to an N-terminal PspA in Actinobacteria and Chloroflexi [Fig. 3; Table S3; e.g., ACU53894.1, Acidimicrobium]. We call this domain PspAA (for PspA-associated domain A; Table S2; web app). In a few bacterial and archaeal lineages, the PspA–PspAA dyad co-occurs with another dyad comprising a membrane-associated Metallopeptidase and a protein with a novel domain, which we termed PspAB (for PspA-associated domain B, AAZ55047.1, Tfu_1009 Thermobifida) [Fig. 3; Table
S2; web app]. This predicted operon occasionally contains a third gene coding for a SHOCT-like bihelical domain-containing protein in various bacterial and archaeal lineages [Fig. 3; Table S3; ABW11964.1, Frankia; AKB54760.1, Methanosarcina; AKX93460.1, Moorella].

2. Novel PSP associations

2.1 PspA/Snf7 domain architectures

A very small fraction of PspA homologs shows variation in their domain architecture (proteins that contain fusions with PspA instead of carrying PspA alone). For example, cyanobacterial PspA homologs show some interesting variations: a few have dyads or triads of PspA, either as repeated domains within a polypeptide or a predicted operon with multiple copies of PspA-containing genes [Fig. 3; e.g., BAG06017.1], while others carry an additional hydrolase domain of NlpC/P60 superfamily at the N-terminus that is predicted to catalyze the modification of phosphatidylcholine, thus altering membrane composition [Fig. 3; Table S3; AFZ52345.1; (1)]. We also find a novel fusion of PspA with PspAA in Actinobacteria (ACU53894.1, Acidimicrobium; defined in the section on PspAA below). Similar to the PspA homologs, a search for the related superfamily, Snf7, revealed minimal variation in domain architecture, with occasional fusions (<5%) found only in eukaryotes [Fig. 2B]. Some Actinobacteria, such as Mycobacteroides abscessus, have an Snf7 homolog [CAM62382.1, Fig. 2B] fused to an RND-family transporter member. The latter transports lipids and fatty acid and is flanked by two genes encoding the Mycobacterium-specific TM protein with a C-terminal Cysteine-rich domain (2).

2.2 Vps4 and AAA⁺-ATPases

One or more copies of an snf7 gene [e.g., OLS27540.1; Table S3] and a gene for the VPS4-like AAA⁺-ATPase (with an N-terminal MIT domain and C-terminal oligomerization domain; Table S2) are known to occur together in Archaea; they define the core of an ESCRT complex (3). However, we observed some diversity between different archaeal lineages. For example, the Asgardarchaeota contain a genomic context that is most similar to eukaryotes. This archaeal context is composed of the Vps4 AAA⁺-ATPase and Snf7-encoding genes along with an ESCRT-II gene that encodes a protein with multiple winged helix-turn-helix (wHTH) domains (4). In Crenarchaeota, Snf7 and the Vps4 AAA⁺-ATPase are encoded in a distinct three-gene operon, which contains a gene coding for a CdvA-like coiled-coil protein with an N-terminal PRC-barrel domain implicated in archaeal cell division (5). In this case, the Snf7 domain is fused to a C-terminal wHTH domain, which might play a role equivalent to the ESCRT-II wHTH domain. These operons may be further extended with additional copies of Snf7 genes and other genes coding for a TM protein and an ABC ATPase. We also observed that a related VPS4-like AAA⁺-ATPase was transferred from Archaea to Bacteria and is found in Cyanobacteria, Bacteroidetes, Verrucomicrobia, Nitrospirae, and Planctomycetes (e.g., ACB74714.1, Opitutus; Table S3). In these operons, the snf7 gene is displaced by an unrelated gene coding for a larger protein with TPR repeats followed by a 6TM domain, again suggesting a membrane-proximal complex.
PSP evolution across the tree of life

Our analysis also showed that the bacterial PspA (e.g., AEY64321.1, Clostridium; Table S3) might occur with a distinct AAA^+^-ATPase in various bacterial clades. The resulting protein (e.g., AEY64320.1, Clostridium) has two AAA^+^-ATPase domains (e.g., CKH37208.1, Mycolicibacterium) in the same polypeptide, with the N-terminal version being inactive. This gene dyad also occurs with either a previously unidentified membrane-anchored protein with a divergent Snf7 domain (OGG56892.1; Table S3) and other coiled-coil or α-helical domain-containing proteins. Both PspA and the membrane-associated Snf7, along with the AAA^+^-ATPase, may occur in longer operons with other genes coding for an ABC-ATPase, an ABC TM permease, and a solute-binding protein with PBPs and OmpA domains [e.g., OGG56892.1; Table S3].

2.3 PspA with PspM or Thioredoxin

The association of ClgR-HTH with PspAM is also confined to this RsmP family, suggesting that these are also determinants of the rod-shaped morphology of the cell. The PspN presence in the immediate operon of ClgR-HTH–PspAM (containing ClgR, PspA, PspM) is limited to a few mycobacteria (CCP45543.1, M. tuberculosis H37Rv), which have an N-terminal PspN_N (as defined below) and C-terminal DUF3046. The remaining ClgR-HTH–PspAM operons lack the fused PspN_N–DUF3046 protein and instead contain only the ancestral DUF3046 located three genes downstream (ABK71106.1, Mycolicibacterium smegmatis). The duplicated DUF3046 domain forms the intact ClgR-HTH–PspAMN operon only in the M. tuberculosis complex (6, 7). The presence of the same family of thioredoxin with a different family of PspA (typically, two copies) in Cyanobacteria suggests that the thioredoxin homolog is involved in a similar redox activity to control PspA [AFZ14666.1, Crinalium; Fig. 3; Table S3].

2.4 Novel contexts containing Toastrack

Toastrack and TM domains

In most homologs, we find that Toastrack domains are fused to N-terminal single or multi-TM domains such as PspC, Lia-LiaF-TM, HAAS, SHOCT, strongly suggesting that the Toastrack domains are predominantly intracellular with N-terminal membrane tethers [Fig. 4, 5; Table S4]. In Cyanobacteria, we find variable multidomain proteins with an N-terminal TM anchor followed by a region containing the Toastrack domain flanked by immunoglobulin (Ig) and one or more catalytic domains such as a fringe-like glycosyltransferase or a caspase-like thiol peptidase [Fig. 4; Table S4; AFY83227.1, Oscillatoria; Table S2]. Further, in several architectures, the N-terminal TM regions fused to the Toastrack domain are replaced by at least two variants of the bihelical SHOCT (e.g., Bacillus subtilis yvlB, CAB15517.1) [Fig. 4, 5; Tables S1 and S4]. We call these variants SHOCT-like domains to distinguish them from the classical SHOCT domain, as these include a domain partly detected by the Pfam DUF1707 (8) model and another that has not been detected by any published profile. The SHOCT and related domains are fused to disparate domains and are typically found at the N- or C-termini of proteins.
We also discovered several conserved genomic contexts containing Toastrack, with likely roles in membrane-linked stress response: The first of these found across diverse bacterial lineages contains a core of four genes coding for i) a sigma factor, ii) a receptor-like single TM protein with an intracellular anti-sigma-factor zinc finger (zf-HC2, PF13490 in Pfam) and extracellular HEAT repeats, iii) one or two membrane-anchored Toastrack-containing proteins (AFK03672.1 Emticicia; Table S4), and iv) a previously uncharacterized protein with hits to the Pfam model DUF2089. We found that this Pfam model DUF2089 can be divided into an N-terminal ZnR, central HTH, and C-terminal SHOCT-like domains (ADE70705.1, Bacillus) [Fig. 4; Tables S2 and S4]. In a few of these operons, the membrane anchor of the Toastrack domain is a LiaI-LiaF-TM domain [Fig. 5; Table S4]. Variants of this system include additional genes coding for a protein with a LiaI-LiaF-TM domain fused to an N-terminal B-box domain (e.g., ACO33311.1, Acidobacterium) or a PspC protein (e.g., OGF50123.1, Candidatus Firestonebacteria) [Fig. 5; Table S4]. We propose that this three-gene system functions similarly to the classical lia operon in transducing membrane-associated signals to a transcriptional output affecting a wide range of genes via the sigma factor.

Similarly, an operon observed predominantly in various Proteobacteria and Bacteroidetes couples a protein with a membrane-anchored Toastrack domain (typified by AAM36414.1, Xanthomonas) with genes coding for an ABC-ATPase, a permease subunit, and a GNTR-HTH transcription factor with distinct C-terminal α-helical domain and another [Fig. 5; Tables S2 and S4]. These operons also code for a previously uncharacterized protein matching the Pfam DUF2884 model. We show that these proteins are membrane-associated lipoproteins (e.g., AJI31452.1, Bacillus), which might function as an extracellular solute-binding partner for the ABC-ATPase and permease components. A comparable operon found in Actinobacteria replaces the GNTR-HTH transcription factor with a ribbon-helix-helix (RHH) domain protein. In some Actinobacteria, the Toastrack domain encoded by the operon is fused to a SHOCT-like domain and is encoded adjacent to genes specifying a two-component system (CCP43715.1, Mycobacterium) or a transport operon (CAB88834.1, Streptomyces) [Fig. 5; Table S4]. These operons with the Toastrack domains are likely to couple transcriptional regulation to the sensing of membrane-proximal signal and transport [Fig. 5; Table S4]. The GNTR-HTH and RHH operons in these systems are likely to function as transcriptional regulators analogous to PspF and ClgR transcription factors from classical PSP systems.
References

1. Anantharaman V, Aravind L. 2003. Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. Genome Biol 4:R11.

2. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eigelmeier K, Garnier T, Gutierrez C, Hewinson G, Kremer K, Parsons LM, Pym AS, Samper S, van Soolingen D, Cole ST. 2002. A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci U S A 99:3684–3689.

3. Wollert T, Wunder C, Lippincott-Schwartz J, Hurley JH. 2009. Membrane scission by the ESCRT-III complex. Nature 458:172–177.

4. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L, Vancaester E, Seitz KW, Anantharaman K, Starnawski P, Kjeldsen KH, Stott MB, Nunoura T, Banfield JF, Schramm A, Baker BJ, Spang A, Ettema TJG. 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature 541:353–358.

5. Kaur G, Iyer LM, Subramanian S, Aravind L. 2018. Evolutionary convergence and divergence in archaeal chromosomal proteins and Chromo-like domains from bacteria and eukaryotes. Sci Rep 8:6196.

6. Datta P, Ravi J, Guerrini V, Chauhan R, Neiditch MB, Shell SS, Fortune SM, Hancioglu B, Igoshin OA, Gennaro ML. 2015. The Psp system of Mycobacterium tuberculosis integrates envelope stress-sensing and envelope-preserving functions. Mol Microbiol 97:408–422.

7. Ravi J, Anantharaman V, Aravind L, Gennaro ML. 2018. Variations on a theme: evolution of the phage-shock-protein system in Actinobacteria. Antonie Van Leeuwenhoek 111:753–760.

8. Eberhardt RY, Bartholdson SJ, Punta M, Bateman A. 2013. The SHOCT domain: a widespread domain under-represented in model organisms. PLoS ONE 8:e57848.