nature microbiology

SUPPLEMENTARY INFORMATION

VOLUME: 2 | ARTICLE NUMBER: 17005

In the format provided by the authors and unedited.

N-terminomics identifies Prli42 as a membrane miniprotein conserved in Firmicutes and critical for stressosome activation in *Listeria monocytogenes*

Francis Impens[1,2,3 #‡], Nathalie Rolhion[1,2,3 #], Lilliana Radoshevich[1,2,3#], Christophe Bécavin[1,2,3,4], Mélodie Duval[1,2,3], Jeffrey Mellin[1,2,3], Francisco García del Portillo[5], M. Graciela Pucciarelli[5,6], Allison H. Williams[7,8], Pascale Cossart[1,2,3,*]

Affiliations

[1]Institut Pasteur, Unité des Interactions Bactéries-Cellules, Département de Biologie Cellulaire et Infection, F-75015 Paris, France.

[2]Inserm, U604, F-75015 Paris, France.

[3]INRA, Unité sous-contrat 2020, F-75015 Paris, France.

[4]Institut Pasteur, Bioinformatics and Biostatistics Hub, C3BI, USR 3756 IP CNRS, Paris, France

[5]Centro Nacional de Biotecnología-Consejo Superior de Investigaciones Científicas (CNB-CSIC), Madrid, Spain.

[6]Departamento de Biología Molecular, Universidad Autónoma de Madrid, Centro de Biología Molecular 'Severo Ochoa' (CBMSO-CSIC), Madrid, Spain

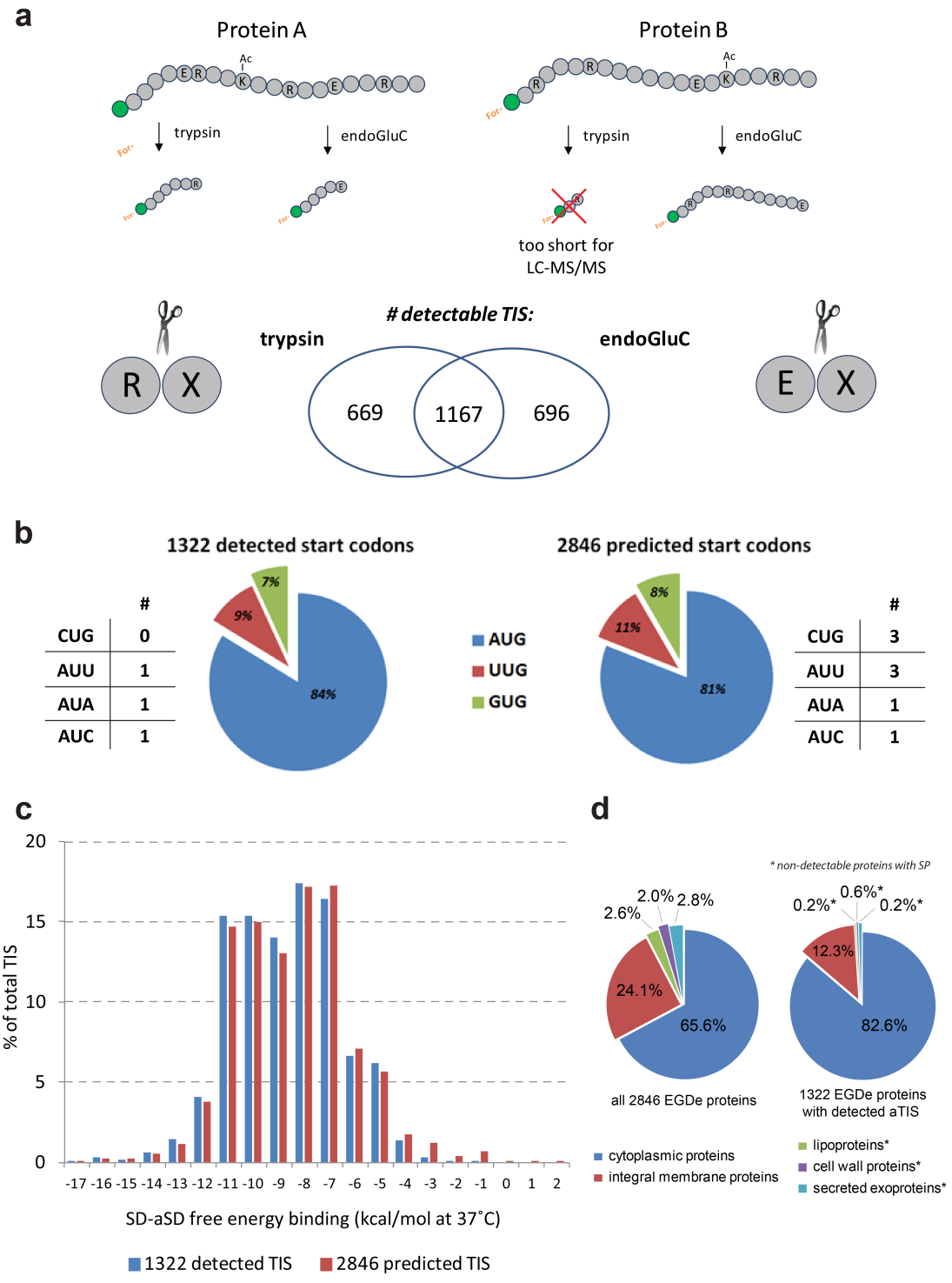
[7]Institut Pasteur, Unité des Biologie et génétique de la paroi bactérienne, Département de Microbiologie, F-75015 Paris, France.

[8]INSERM, Groupe Avenir, F-75015 Paris, France.

‡ Present address: Medical Biotechnology Center, VIB, Ghent University, 9000 Ghent, Belgium.
# These authors contributed equally to this work.

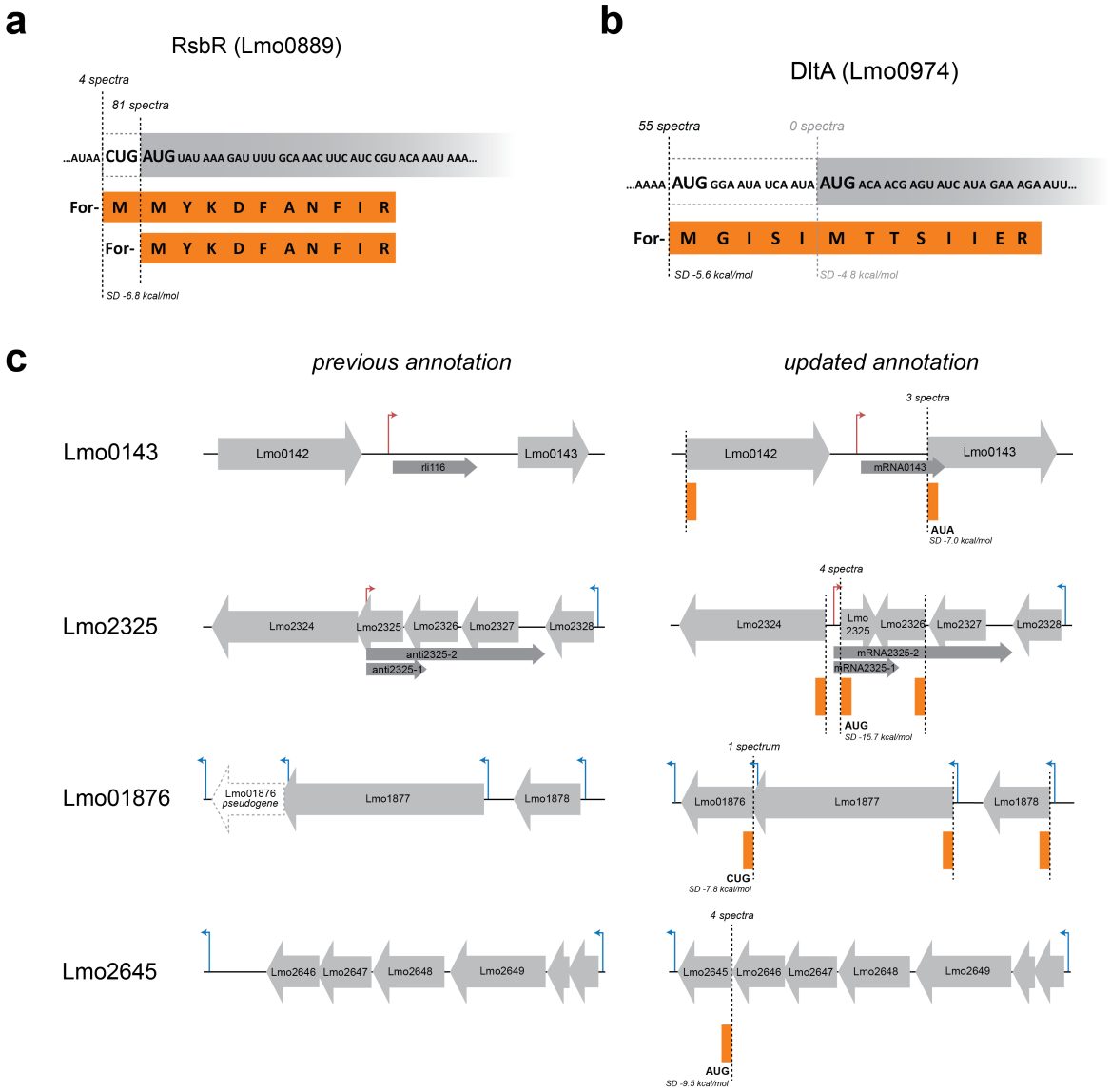
** Correspondence to: pascale.cossart@pasteur.fr*

Supplementary Figures

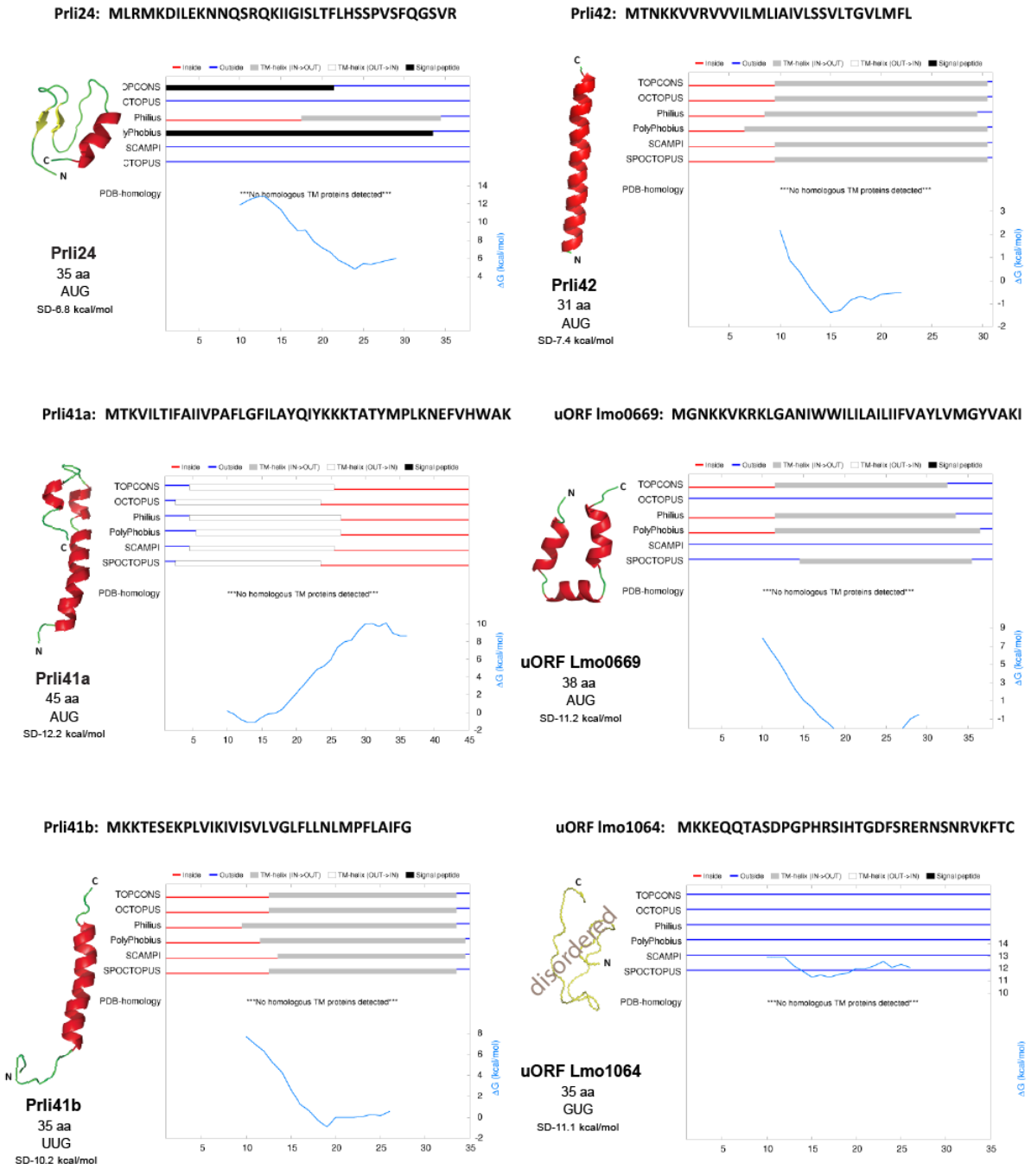


Supplementary Figure 1. Extended results of N-terminal COFRADIC (a) Trypsin and endoproteinase GluC (endoGluC) have complementary coverage for the *Listeria* N-terminome. Different enzymatic digestions release different N-terminal peptides depending on the enzyme specificity. Trypsin cleaves after Arg and Lys residues,

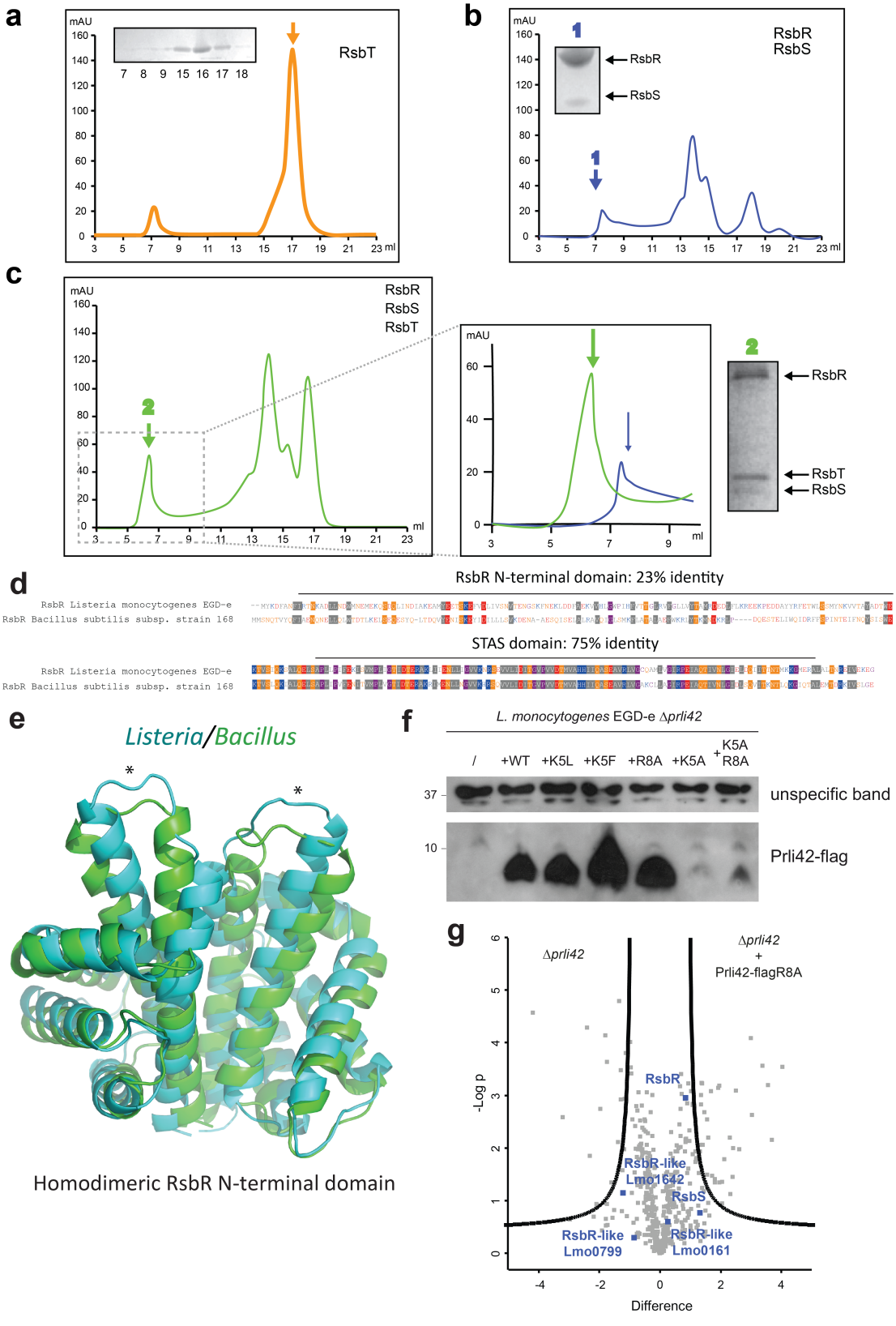
but since the latter are acetylated during sample preparation, trypsin only cleaves after Arg residues in the N-terminal COFRADIC protocol. In contrast, endoGluC cleaves after acidic residues, primarily Glu. The position of the first Arg or Glu residue in a protein will thus determine the length and the mass of the N-terminal peptide upon trypsin or endoGluC cleavage, respectively. **(b)** Pie charts showing the distribution of AUG, UUG and GUG start codons for all 2846 predicted aTIS (annotated Translation Initiation Sites) (right) and the subset of 1322 detected aTIS (left). The number of non-canonical start codons is indicated for each category below. **(c)** The SD-aSD (obtained Shine-Dalgarno-anti Shine-Dalgarno) free binding energy was calculated for every TIS as described under Materials and Methods. The distribution of the calculated values is shown for all 2846 predicted aTIS (red) and the subset of 1322 detected aTIS (blue) as a percentage of the total number of TIS in each category. **(d)** Predicted subcellular localization of proteins with detected aTIS. Pie chart showing the distribution of the predicted subcellular localization for all predicted EGD-e proteins (left) and for 1322 proteins with detected aTIS (right).



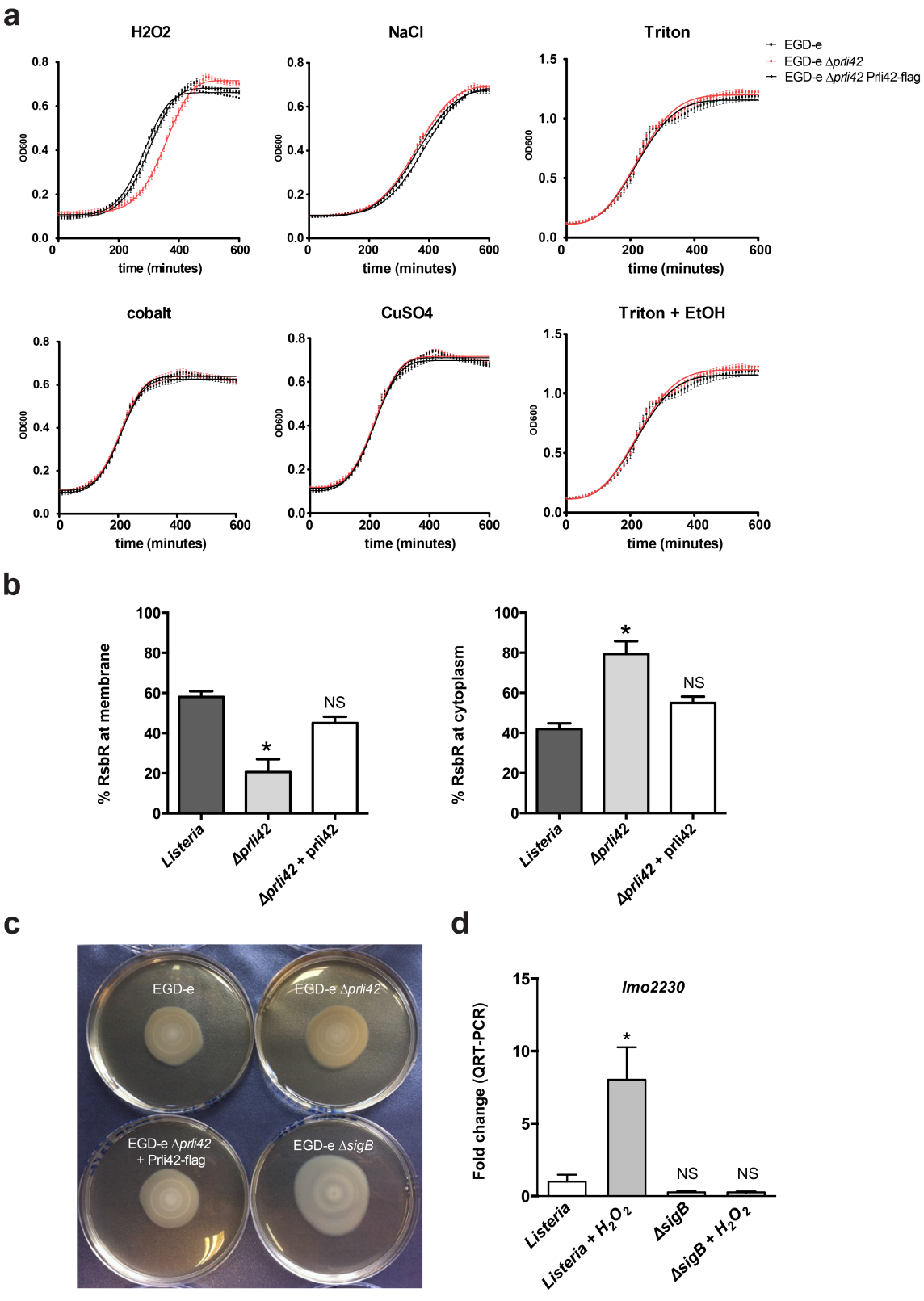
Supplementary Figure 2. Small deviations from the annotated genome detected by our approach. (a) Multiple TISs were detected for RsbR. **(b)** In the case of DltA, no N-terminal peptides corresponding to the annotated TIS were found, 55 TIS peptides were from an upstream AUG start codon. **(c)** Evidence for mis-annotated or missing EGD-e proteins. The previous annotation of the genome loci surrounding the genes encoding Lmo0143, Lmo01876, Lmo2325 and Lmo2645 is shown on the left. On the right we show the N-terminal peptides that mapped to these loci (Supplementary Table 6).



Supplementary Figure 3. Prediction of membrane protein 3D structure and topology. The predicted topology and predicted ΔG values from the TOPCONS web server (http://topcons.cbr.su.se/) are shown for all six *Listeria* miniproteins (Supplementary Table 8). The three-dimensional structure was predicted by molecular modeling and is shown for each protein together with the amino acid length, start codon and the hybridization free energy against the *Listeria* consensus anti-Shine-Dalgarno (aSD) sequence determined as described under Materials and Methods.


a
mAU
RsbT
7 8 9 15 16 17 18
ml
b
mAU
RsbR
RsbS
1
RsbR
RsbS
ml
c
mAU
RsbR
RsbS
RsbT
2
ml
2
RsbR
RsbT
RsbS
d
RsbR N-terminal domain: 23% identity
RsbR Listeria monocytogenes EGD-e
RsbR Bacillus subtilis subsp. strain 168
STAS domain: 75% identity
RsbR Listeria monocytogenes EGD-e
RsbR Bacillus subtilis subsp. strain 168
e
Listeria/Bacillus
Homodimeric RsbR N-terminal domain
f
L. monocytogenes EGD-e Δprli42
/ +WT +K5L +K5F +R8A +K5A +K5A R8A
37
unspecific band
10
Prli42-flag
g
Δprli42
Δprli42 + Prli42-flagR8A
-Log p
RsbR
RsbR-like Lmo1642
RsbS
RsbR-like Lmo0799
RsbR-like Lmo0161
Difference

Supplementary Figure 4. Reconstitution of the *Listeria* stressosome and interaction between Prli42 and RsbR (a) Full trace of the gel filtration chromatograms (UV absorption at 280 nm) of the purified RsbT protein, **(b)** the RsbR:RsbS and **(c)** the RsbR:RsbS:RsbT stressosome complexes shown in Figure 3c. **(d)** N-terminal sequence alignment of *Listeria* and *Bacillus* RsbR. **(e)** Overlay of the crystal structure of the homodimeric RsbR N-terminal domain from *Bacillus* (light green) (PDB ID: 2BNL) with the predicted structure of *Listeria* RsbR (blue). **(f)** Evaluation of the stability of Prli42-flag variants. Complementation of *L. monocytogenes* EGD-e deficient for Prli42 with C-terminally flag-tagged substitution variants of the miniprotein by immunoblot. **(g)** Role of R8A in the binding to stressosome related proteins. Prli42-R8A-flag was expressed in the *rli42* deletion strain, immunoprecipitated and co-immunoprecipitated proteins were identified by LC-MS/MS. For every protein, the volcano plot shows the intensity fold change (in $\log_2$) in the Prli42-R8A expressing strain compared to the control deletion strain on the X-axis. The pull down was performed in triplicate and a t-test was performed to calculate -log p values for each protein indicated on the Y-axis, black lines indicate the boundery of significance as set by the Perseus software (FDR=0.05 and S0=1). Elution profiles and immunoblots shown in a-c and f are representative of three independent experiments.


a
H2O2
NaCl
Triton
cobalt
CuSO4
Triton + EtOH
OD600
time (minutes)
EGD-e
EGD-e Δprli42
EGD-e Δprli42 Prli42-flag
b
% RsbR at membrane
% RsbR at cytoplasm
NS
*
Listeria
Δprli42
Δprli42 + prli42
c
EGD-e
EGD-e Δprli42
EGD-e Δprli42 + Prli42-flag
EGD-e ΔsigB
d
Imo2230
Fold change (QRT-PCR)
Listeria
Listeria + H2O2
ΔsigB
ΔsigB + H2O2

Supplementary Figure 5. Effect of different stresses on bacterial growth, membrane versus cytoplasmic localization of RsbR and validation that Sigma B activates lmo2230 following oxidative stress. (**a**) Time course of *Listeria* strains growth in BHI + 0.05% H_2O_2, 4% NaCl, 200uM $CoSO_4$, 400uM $CuSo_4$, 0.025% Triton or 0.025% Triton and 2% Ethanol. Results are means ± SEM of 3 independent experiments. (**b**) Quantification of percentage of RsbR in membrane or cytosoplasmic fractions using ImageJ. Values from membrane and cytoplasmic RsbR were totaled and percentages were determined for each of the three strains (WT, *Δprli42* and *Δprli42* + Prli42-flag). Results are expressed as mean ± SEM of three independent experiments. (**c**) Ring forming abilities on an agar plate at room temperature. Bacteria were inoculated on low-agar plates and exposed to five cycles of 12h light/12h darkness. Image is representative of three independent experiments. (**d**) Fold change was calculated using the ΔΔCT method for *lmo2230* relative to *Listeria rpoB*. Results are expressed as mean ± SEM of at least three independent experiments. For all data significance was determined by ordinary one-way ANOVA followed by Dunnett's multiple comparison test, significance is shown relative to WT in each condition (*** $p < 0.001$, * $p < 0.05$, NS = not significant).

Figure 2C

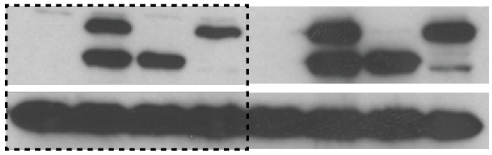


Figure 2D

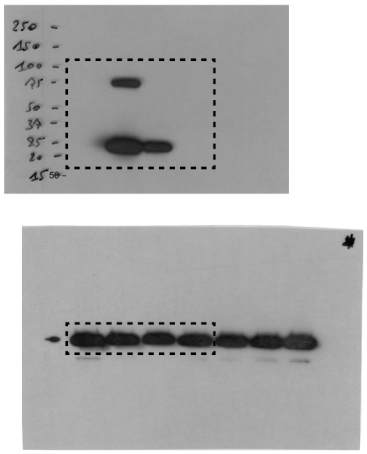


Figure 3F (top)

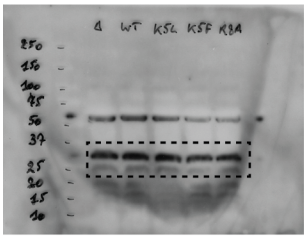


Figure 3F (top)

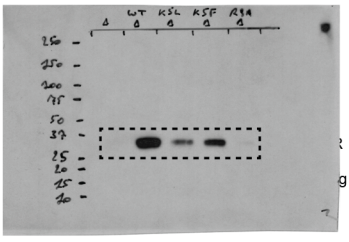
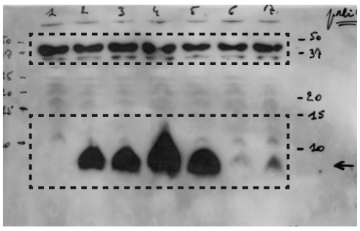
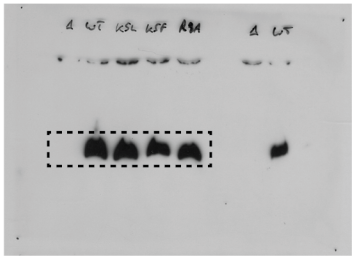
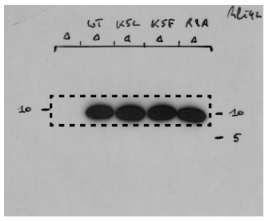


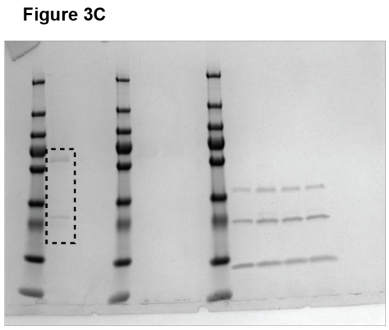
Figure S4F

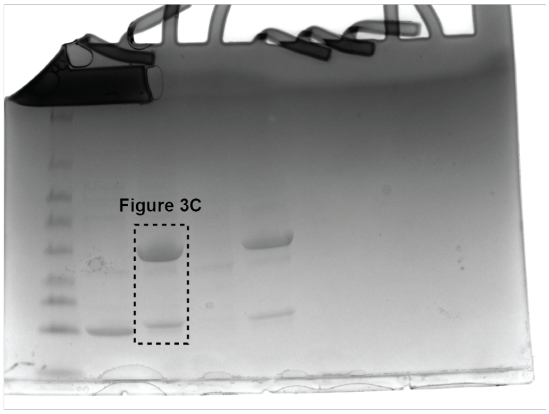






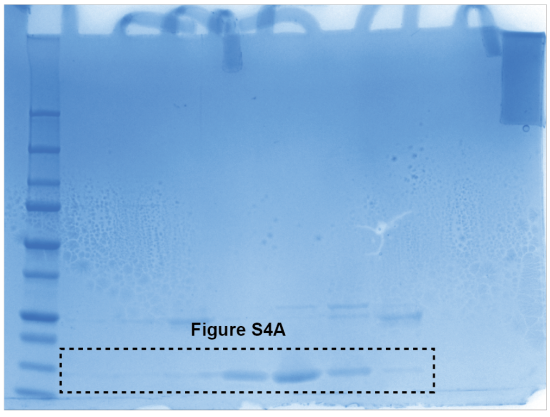
Supplementary Figure 6. Full blots from figures.





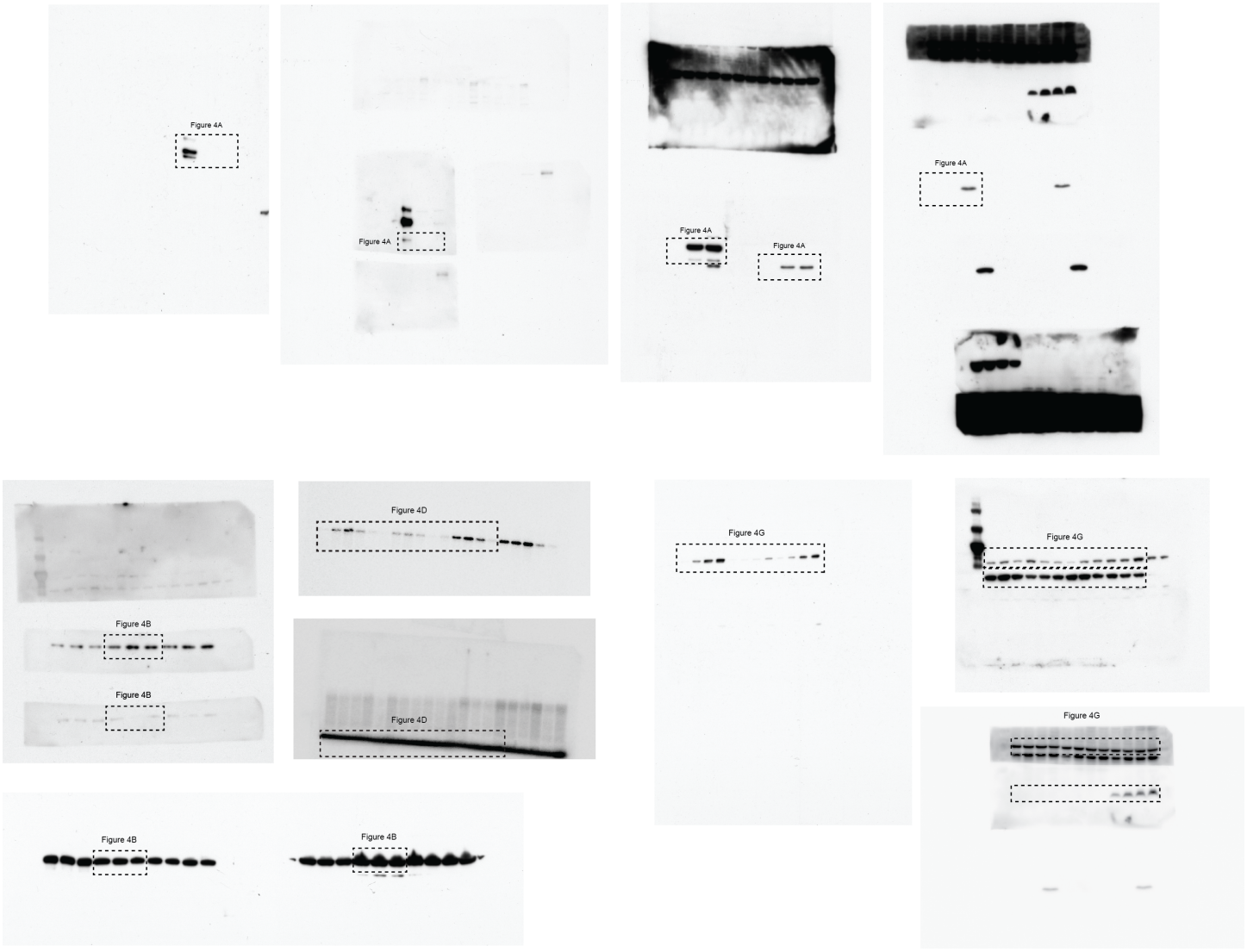
Location: C:/Users/Geldoc XR/Desktop/USERS/Constance
Printed: 08/03/2016 12:20:00 Page 1 of 1

Geldoc XR 2016-03-02 18hr 04min



Printed: 02/03/2016 18:06:54 Page 1 of 1

Supplementary Figure 7. Full blots from figures.



Supplementary Figure 8. Full blots from figures.

Supplementary Tables:

Supplementary Table 1. **List of predicted undetectable aTIS in *L. monocytogenes* EGD-e.** Columns from left to right indicate whether the protein is predicted to be non-expressed (sheet 1), to be secreted extracellular and/or have an N-terminal signal peptide removed in their mature form (sheet 2) or to possess an MS undetectable N-terminal peptide released by trypsin or endoGluC (sheet 3).

Supplementary Table 2. List of 1322 *L. monocytogenes* EGD-e proteins with detected annotated TIS (aTIS). Columns from left to right contain the protein (lmo) number, the RAST protein (RAST_lmo) number, the gene name, the start codon, the SD base-pairing free energy in kcal/mol, the strand of the genome on which the gene is present, the genome start position of the ORF, the genome start position of ORF predicted by RAST, the total number of MS/MS spectra of the corresponding N-terminal peptides, the sequence of the best scoring (For-; formylated) N-terminal peptide with its Mascot score and threshold, an indication whether the peptide was found upon endoGluC or trypsin digestion, a description of the protein function and the COG functional annotation of the protein.

Supplementary Table 3. List of 72 *L. monocytogenes* EGD-e proteins with leaderless mRNAs. Thirteen proteins with only evidence for a single leaderless transcript are shown on top. Columns from left to right contain the protein (lmo) number, the gene name, the start codon, the SD base-pairing free energy in kcal/mol, the strand of the genome on which the gene is present, the genome position of the TIS, the genome position of the nearest TSS, the distance in bp between both, the complete TSS counts (total counts over all conditions tested) and an indication whether other TSS are present as well.

Supplementary Table 4. List of 27 *L. monocytogenes* EGD-e proteins with multiple TIS. Columns from left to right contain the N-terminal truncated or extended sequence compared to the annotated sequence, the protein (lmo) number, the gene name, description and COG annotation, the strand of the genome on which the gene is present and the genome start position of the ORF. The following columns contain information about the detection of the annotated TIS (aTIS) and the alternative TIS. They list the start position, start codon, the SD base-pairing free energy in kcal/mol, the total number of MS/MS spectra, the sequence of the best scoring (For-; formylated) N-terminal peptide with its Mascot score and threshold and an indication whether the alternative TIS has been predicted by RAST.

Supplementary Table 5. List of 25 *L. monocytogenes* EGD-e proteins with a corrected TIS. These are proteins for which the annotated TIS was never detected (although detectable), instead only a nearby alternative TIS was picked up. Columns from left to right contain the N-terminal truncated or extended sequence compared to the annotated sequence, the protein (lmo) number, the gene name, description and COG annotation, the strand of the genome on which the gene is present. The following columns contain information about the detection of the annotated TIS (aTIS) and corrected TIS and list the start position of the annotated ORF, start codon, the SD base-pairing free energy in kcal/mol, the total number of MS/MS spectra, the sequence of the best scoring (For-; formylated) N-terminal peptide with its Mascot score and threshold and an indication whether the corrected TIS is predicted by RAST.

Supplementary Table 6. List of two mis-annotated and two missing *L. monocytogenes* EGD-e proteins. Columns from left to right contain the protein (lmo) number, description and COG annotation, the strand of the genome on which the gene is present, the genome start position of the ORF, the start codon, the SD base-pairing free energy in kcal/mol, the total number of MS/MS spectra, the sequence of the best scoring (For-; formylated) N-terminal peptide with its Mascot score and threshold, the corrected amino acid protein length and sequence and a comment whether the protein was mis-annotated or missing in the original annotation and/or predicted by RAST.

Supplementary Table 7. List of 19 *L. monocytogenes* EGD-e proteins with detected internal TIS. Columns from left to right contain the protein (lmo) number, the gene name, description and COG annotation and the strand of the genome on which the gene is present. The following columns contain information about the detection of the annotated TIS (aTIS) and the internal TIS (iTIS). They list the genome start position, the start codon, the SD base-pairing free energy in kcal/mol, the total number of MS/MS spectra and the sequence of the best scoring (For-; formylated) N-terminal peptide with its Mascot score and threshold.

Supplementary Table 8. List of six newly identified miniproteins in *L. monocytogenes* EGD-e. Columns from left to right contain the protein name and description, the strand of the genome on which the gene is present, the genome start position and start codon of the ORF, the SD base-pairing free energy in kcal/mol, the total number of MS/MS spectra by which the ORF was detected, the sequence of the best scoring (For-; formylated) N-terminal peptide with its Mascot score and threshold, the amino acid protein length and sequence, an indication whether the peptide was also identified by MaxQuant and a comment why the selected ORF was annotated if no For-Met starting N-terminal peptide was identified or if the peptide was not identified by MaxQuant.

Supplementary Table 9. Results of the homolog search of Rli42, and the stressosome. The first sheet, contains all the BlastP results of Rli42, RsbR, RsbS, RsbT, the NTerm globin domain of RsbR and the STAS domain of RsbR searched against the 5000 complete bacterial genomes available on RefSeq database. The other sheets display the list of paralogs of RsbR, RsbS and RsbT found for each bacterial genome.

Supplementary Table 10. List of *L. monocytogenes* EGD-e proteins identified and quantified by LC-MS/MS after co-immunoprecipitation of Prli42-flag or Prli42-R8A-flag. Columns from left to right contain the MaxQuant majority protein ID, difference and -log p value calculated by the Perseus t-test used to compare both samples, an indication whether the protein was identified as significant hit, the LFQ intensity values in each replicate sample, the total intensity value, the MaxQuant PEP score, an indication whether the protein was identified as potential contaminant and the MaxQuant protein ID.

Supplementary Table 11. Results of the RAST re-annotation of the *L. monocytogenes* EGD-e genome and comparison with the original annotation by Glaser et al (referred to as NCBI). The different sheets contain information on the ORFs that were predicted in both annotations (exact matches), ORFs for which a different start codon was predicted at a distance of less than 100 bp, ORFs for which a different start codon was predicted at a distance of more than 100 bp, ORFs that were predicted as pseudogenes in the original annotation and ORFs that were predicted only by RAST or in the original annotation (unique matches).

Supplementary Table 12. Strains, plasmids and primers used in this study.

Supplementary Table 13. Spectral counts and peptide numbers for the different datasets. The number of recorded, identified and removed spectra as well as the number of N-terminally blocked and N-terminally formylated peptides is indicated for each dataset.