

Signatures of Conformational Stability and Oxidation Resistance in Proteomes of Pathogenic Bacteria

Anita Vidovic,¹ Fran Supek,^{2,3,4} Andrea Nikolic,¹ and Anita Krisko^{1,*}

¹Mediterranean Institute for Life Sciences, Mestrovicevo setaliste 45, 21000 Split, Croatia

²Division of Electronics, Rudjer Boskovic Institute, Bijenicka cesta 54, 10000 Zagreb, Croatia

³EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain

⁴Universitat Pompeu Fabra (UPF), 08002 Barcelona, Spain

*Correspondence: anita.krisko@medils.hr

<http://dx.doi.org/10.1016/j.celrep.2014.04.057>

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

SUMMARY

Protein oxidation is known to compromise vital cellular functions. Therefore, invading pathogenic bacteria must resist damage inflicted by host defenses via reactive oxygen species. Using comparative genomics and experimental approaches, we provide multiple lines of evidence that proteins from pathogenic bacteria have acquired resistance to oxidative stress by an increased conformational stability. Representative pathogens exhibited higher survival upon HSP90 inhibition and a less-oxidation-prone proteome. A proteome signature of the 46 pathogenic bacteria encompasses 14 physicochemical features related to increasing protein conformational stability. By purifying ten representative proteins, we demonstrate in vitro that proteins with a pathogen-like signature are more resistant to oxidative stress as a consequence of their increased conformational stability. A compositional signature of the pathogens' proteomes allowed the design of protein fragments more resilient to both unfolding and carbonylation, validating the relationship between conformational stability and oxidability with implications for synthetic biology and antimicrobial strategies.

INTRODUCTION

As a part of their strategy against pathogen invasion, host organisms expose pathogens to deteriorating effects of reactive oxygen species (ROS) (Rosenberger and Finlay, 2003). In such environments, survival of invading bacteria depends on their tolerance of the oxidative stress. Consequently, microorganisms have evolved several known strategies for defense against ROS; while the activity of antioxidant enzymes scavenges ROS, a decreased amount of iron-sulfur clusters in proteins lowers the toxicity resulting from their destruction (Imlay, 2003).

Accumulation of oxidative damage in proteins, most importantly protein carbonylation, can cause irreversible and progressive degeneration of almost all cellular functions (Krisko and Radman, 2013), resulting in cell death (Krisko and Radman, 2010). Three factors contribute to total protein carbonylation: (1) the rate of ROS production, (2) the intrinsic resistance of individual proteins to oxidative damage, and (3) the damaged proteins' clearance rate by degradation and/or aggregation (Nyström, 2005). Here, we explore the details of the inequalities in the proteins' susceptibility to oxidative damage, when examined at a larger, whole-proteome scale. Even though various protein folds may differ in their resistance to oxidative stress, natively folded proteins are known to be, by and large, more robust than misfolded ones (Dukan et al., 2000).

The availability of thousands of complete prokaryotic proteome sequences allows in-depth comparisons of their global properties determined from single-residue compositions (Tekaiia and Yeramian, 2006). So far, such in silico studies have mainly focused on adaptation of proteomes to extreme environments (Zeldovich et al., 2007; Smole et al., 2011), while a global characterization of proteomic signatures of the more prevalent microbial lifestyles, such as commensalism or pathogenicity, has received less attention. In the present study, we investigate the existence and nature of the common properties in pathogen proteomes and the phenotypic consequences thereof.

In particular, the high oxidation sensitivity of misfolded proteins (Dukan et al., 2000) led us to hypothesize a greater intrinsic resistance to misfolding of proteins from pathogenic bacteria, which could provide resilience against ROS-based host defenses. This rationale is further supported by the observation that pathogens tend to have lower genomic G+C content, known to favor the incorporation of hydrophobic amino acids, rendering the proteome more resistant to unfolding (Mendez et al., 2010).

We demonstrate in vivo that representative pathogens display an increased resistance to HSP90 chaperone inhibition, paralleled with an increased proteome resistance to in vitro oxidation. The salient trends in the proteomes' physicochemical properties readily distinguished the pathogens from nonpathogens, encoding a signature of the pathogenic proteome characterized by an avoidance of secondary-structure-destabilizing amino acids. Finally, we provide experimental evidence that the pathogens'

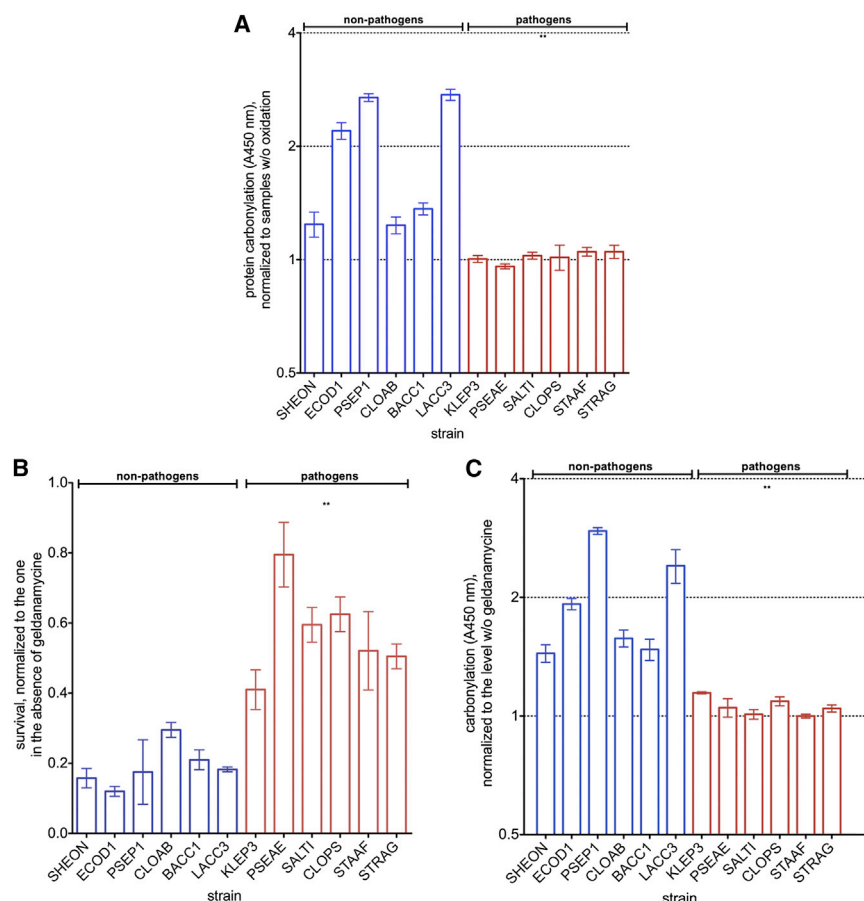


Figure 1. Oxidation Resistance and HSP90 Inhibition in Pathogens and Nonpathogens

(A) Oxidation resistance of total protein extracts, (B) survival, and (C) protein carbonylation in SHEON, ECOD1, PSEP1, CLOAB, BACC1, LACC3, KLEP3, PSEAE, SALT1, CLOPS, STAAF, and STRAG in the presence of 2 μ M geldanamycin. Oxidation resistance in (A) is expressed as protein carbonylation of the oxidized extract relative to the nonoxidized one. The error bars represent the SD of three triplicate measurements. ** $p < 0.01$ (two-tailed t test).

The same amount of geldanamycin entered into the cells of all studied bacterial species, ruling out differential membrane permeability as a cause for the low susceptibility of pathogens to geldanamycin (Supplemental Results; Figure S1A). The presence of geldanamycin did not cause an increase in ROS production in any of the tested species (Supplemental Results; Figures S1B and S1C), excluding the possibility that the pathogens' resistance to geldanamycin relies on efficient suppression of geldanamycin-induced ROS. Rather, the data suggest the existence of a relationship between an increased protein conformational stability and resistance to oxidation in the 12 selected bacteria.

proteins are indeed more resistant to oxidation as a direct consequence of their increased conformational stability.

RESULTS

Pathogen Proteomes Are More Resistant to Oxidation In Vitro and to HSP90 Inhibition In Vivo

We selected six representative pathogenic and six nonpathogenic bacteria and tested their total protein extracts for susceptibility to oxidation. We found that protein extracts from representative pathogens were on average 1.9-fold more resistant to carbonylation than the extracts from nonpathogens ($p = 0.0002$) (Figure 1A; species names are listed in Table S1).

Furthermore, we measured the resistance of the 12 bacteria to geldanamycin, a known inhibitor of the HSP90 chaperone. An intrinsically stable proteome should be able to withstand a HSP90 deficiency in normal conditions, reflected in high cell survival and a low level of geldanamycin-induced protein damage. Indeed, geldanamycin exposure resulted in 19% average survival for the nonpathogens (Figure 1B), in contrast to 58% for the pathogens ($p < 0.0001$). The geldanamycin-induced mortality was paralleled by the total protein carbonylation levels (Figure 1C); lower mortality of the pathogenic bacteria was correlated with a 1.9-times-lower average level of protein carbonylation relative to nonpathogens ($p < 0.0001$).

Furthermore, we evaluated the protection against geldanamycin granted by the antioxidant enzymes KatE, KatG, and SodD and by the small-molecule antioxidant N-acetylcysteine (NAC) in *E. coli*. Regardless of the source of the antioxidant activity, it can account for, on average, only 18% of the difference in geldanamycin resistance observed between the pathogens and nonpathogens reported in Figure 1B (Supplemental Results; Figure S1D). Namely, while the four sources of antioxidant activity yielded an average 30% increase in survival in the presence of geldanamycin, the average survival difference between the six pathogenic and six nonpathogenic bacteria was 3.0-fold (Figure 1B). Consistent trends were observed in the level of protein carbonylation (Supplemental Results; Figure S1E). In other words, the extent of proteome oxidation protection that can be obtained by increased amounts of antioxidants appears to be small in comparison to the differences we observed between pathogenic and nonpathogenic bacteria. Thus, such differences likely also stem from other sources, such as intrinsic properties of the proteome.

Pathogenicity Is Reflected in the General Trends of Variation between Proteome Physicochemical Properties

Next, we investigated whether the resilience of the six chosen pathogens to HSP90 inhibition and proteome oxidation is representative of a broader trend across more microbes. We thus

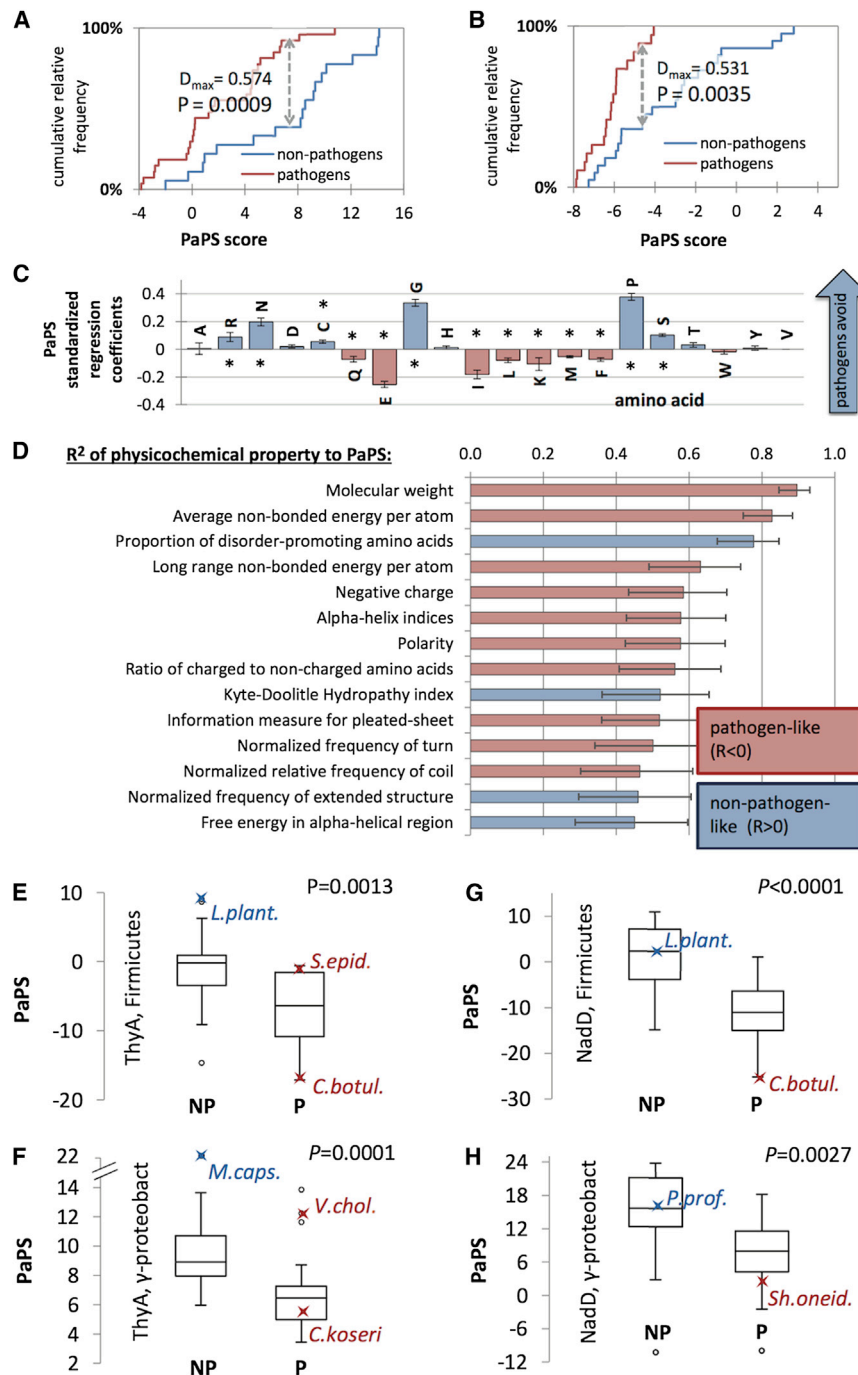


Figure 2. Dominant Trends in the Proteomes' Physicochemical Properties

(A and B) Cumulative histograms of PaPS scores for the pathogens and nonpathogens. Pathogen proteomes tend toward lower PaPS. D_{\max} is the Kolmogorov-Smirnov (K-S) statistic and p value is by K-S test (two tailed).

(C) Standardized coefficients from multiple linear regression of the PaPS signature against amino acid frequencies. $*p < 0.001$ (t test).

(D) The strongest correlations between PaPS and a nonredundant set of amino acid physicochemical properties.

Error bars in (C) and (D) are 95% confidence intervals.

(E–H) Distributions of PaPS scores within ThyA and NadD protein families, contrasted between pathogens and nonpathogens; p values are by two-tailed Mann-Whitney test. The proteins selected for subsequent experiments are highlighted and the abbreviated species names shown. An outlier (PaPS = -10.4) in the set of pathogens in (F) is omitted for clarity.

2B; $p = 0.0035$ and 0.0009 , two-tailed Kolmogorov-Smirnov test). Crucially, we observed a strong agreement between the direction of the pathogenicity-related trend in Firmicutes (Figure S2A) and in γ -proteobacteria (Figure S2B), which is suggestive of a convergent evolution toward a common pathogen-like proteome composition.

Next, we combined the first two PCs into the summary statistic PaPS (the pathogen-like proteome signature) that maximizes the separation between the pathogen and nonpathogen proteomes (Experimental Procedures); pathogens have significantly lower PaPS values (Figures 2A and 2B). Upon examining the contribution of individual amino acid frequencies to PaPS by multiple regression, we found strongly significant contributions of the secondary-structure breakers Gly and Pro (Figure 2C; $p < 0.0001$ for both) followed by Arg and Ser ($p < 0.0001$ for both), all of which are known to have disorder-promoting effects (Uversky, 2013). The avoidance of secondary-

structure-disfavoring amino acids in the pathogens' proteomes suggests a general trend toward more structured proteins. An examination of the physicochemical properties strongly correlated with PaPS (Figure 2D) indicated that pathogens favor bulkier, polar, and charged residues (generally having high β sheet propensities and stabilizing effects) but disfavor residues with high propensity for extended structures. In addition, pathogen proteomes tend toward a lower free energy of α helices. A direct comparison of amino acid propensities between the

compared proteomes of 46 known pathogenic to 40 matched nonpathogenic bacteria according to their overall physicochemical properties, calculated from the genome-encoded amino acid composition (Supplemental Experimental Procedures). In a principal component (PC) analysis, which captures the salient differences between bacterial proteome composition (73% variability in first two PCs; Figure S2), the pathogen proteomes were significantly shifted from the nonpathogens, and the difference was observed separately in two bacterial phyla (Figures 2A and

pathogens and nonpathogens is broadly consistent with the above, with an avoidance of Gly and Pro in pathogens (Figures S2C–S2E; $p \leq 0.05$ and $p \leq 0.016$, respectively, Mann-Whitney test) and an enrichment for the order-promoting Phe and Tyr (Figures S2C–S2E; $p \leq 0.045$ and $p \leq 0.006$, respectively).

While the PaPS scores were, overall, highly significantly different between the pathogens and nonpathogens (Figures 2A and 2B), there were also exceptions to this trend. Such cases may reflect the subtleties in the definition of pathogenicity that are difficult to capture using a binary classification. For instance, two lowly virulent pathogens clustered with nonpathogens, and two insect-only pathogens, clustered with the vertebrate pathogens (Supplemental Discussion; organisms highlighted in Figure S2). Moreover, two nonpathogenic *Listeria* cluster tightly with a close pathogenic relative that they have recently diverged from (Supplemental Discussion).

The observed proteome signature of elevated conformational stability in pathogens could theoretically result from other adaptations (for instance, to the hosts' body temperatures). However, a regression of the organisms' PaPS scores against seven bacterial phenotypes indicates specifically the pathogenicity as the strongest contributor to PaPS, as opposed to, e.g., the optimal growth temperature (Figures S3A and S3B; phenotypes listed in Table S2). Moreover, a known thermophilic proteome signature (Zeldovich et al., 2007) does not strongly relate to PaPS or to the composition of pathogen proteomes (Figures S3C and S3D). Furthermore, we turn to experimentally evaluate whether differences exist between the representative pathogens and nonpathogens in their ability to withstand temperature increases. Although slight differences in survival can be observed between the studied species, they cannot be attributed to their pathogenicity (Supplemental Results; Figures S3E and S3F).

To experimentally validate the trends across proteomes captured by the PaPS statistic, we investigated a series of ten purified proteins selected from both pathogenicity classes and both phylogenetic groups of bacteria (Supplemental Experimental Procedures). We chose the thymidylate synthase (ThyA) and the nicotinic acid mononucleotide adenylyltransferase protein (NadD) families, which exhibited highly significant differences of PaPS scores between the pathogens and nonpathogens (Figures 2E–2H; for NadD, $p < 0.0001$ and $p = 0.0027$ for Firmicutes and γ -proteobacteria; for ThyA, $p = 0.0013$ and 0.0001). Thus, these two protein families are representative of the broader across-proteome trends resulting from the PC analysis.

An Increase in Conformational Stability Is Coupled to an Increase in Oxidation Resistance

Within the NadD and ThyA protein families, we selected six and four proteins, respectively, with highly pathogen-like or highly non-pathogen-like PaPS values (highlighted in Figures 2E–H). These proteins were produced in *E. coli* to experimentally corroborate the link of the PaPS signature to conformational stability and oxidation resistance. By measuring protein carbonylation kinetics, we found that ThyA and NadD proteins from pathogens were significantly more resistant to oxidation relative to their homologs from nonpathogens (Figures 3A and 3B). In particular, the ThyA and NadD with non-pathogen-like PaPS

were 68% and 40% more carbonylated ($p = 0.0004$ and 0.0005), respectively, after Fenton-reaction-induced oxidation relative to the pathogen-like ThyA and NadD. Consistently, ThyA from control pathogens (atypical pathogen proteins that have non-pathogen-like PaPS) oxidized similarly to the ones from nonpathogenic bacteria. This demonstrates that it is specifically the PaPS signature and not some other feature of the pathogens' proteome composition that leads to the oxidation resistance.

Furthermore, to determine their conformational stability and its potential correlation to oxidation resistance, we measured the absorbance at 280 nm during thermal denaturation of ThyA and NadD. Purified ThyA proteins from the two pathogenic bacteria had a melting temperature (T_m) on average 23°C higher than the nonpathogen ThyA (Figures 3C and S4A). Similarly, NadD showed a 21°C average difference (Figures 3C and S4B) ($p = 0.0001$). On the contrary, ThyA proteins from the negative control pathogens (pathogen proteins with a non-pathogen-like PaPS; see Experimental Procedures) had a T_m on average 26°C lower than their homologs from pathogens, thus resembling ThyA from the nonpathogenic bacteria. The large differences in the melting temperature indicate a highly increased conformational stability of the proteins with a pathogen-like PaPS signature. We further tested the link between the two variables by directly modifying the conformational stability using trimethylamine N-oxide (TMAO), an osmolyte known to stabilize protein structures.

Exposure of ThyA and NadD from nonpathogenic bacteria to 1 M TMAO increased the T_m of these proteins by 15°C on average ($p = 0.011$), indicating an increase in conformational stability (Figure 3C; Table S3). Such TMAO-stabilized ThyA and NadD also displayed an increased resistance to oxidation in vitro (Figure 3B); the levels of protein carbonylation were on average 37% lower for ThyA and 23% lower for NadD ($p < 0.0001$ for both; Figure 3D). The control ThyA proteins (proteins from pathogens characterized by non-pathogen-like PaPS) showed similar trends (Figure 3D), with an average decrease of protein carbonylation of 24% following TMAO stabilization ($p < 0.0001$). Furthermore, while stabilization of the proteins makes them more oxidation resistant, the converse also holds; denaturation of all tested proteins by guanidinium hydrochloride (GuHCl) rendered them all extremely sensitive to oxidation (Figure 3E; 2.7-fold average carbonylation increase, $p < 0.0001$). The experiments with external protein stabilization (via TMAO) or destabilization (via GuHCl) corroborate the hypothesis that the protective role of native protein structure against oxidation stems specifically from conformational stability and not from an unrelated property of structure and/or sequence composition.

The PaPS Signature Is Sufficient to Design Conformationally Stable and Oxidation-Resistant Peptides

We aimed to further establish a causal link between the specific proteome signature recognized herein, PaPS, and the detected oxidation resistance and conformational stability. To this end, we have designed a series of six helix-turn-helix (HTH) and six β -hairpin (BH) peptides with amino acid composition selected according to PaPS (see Experimental Procedures; peptides are listed in Table S4).

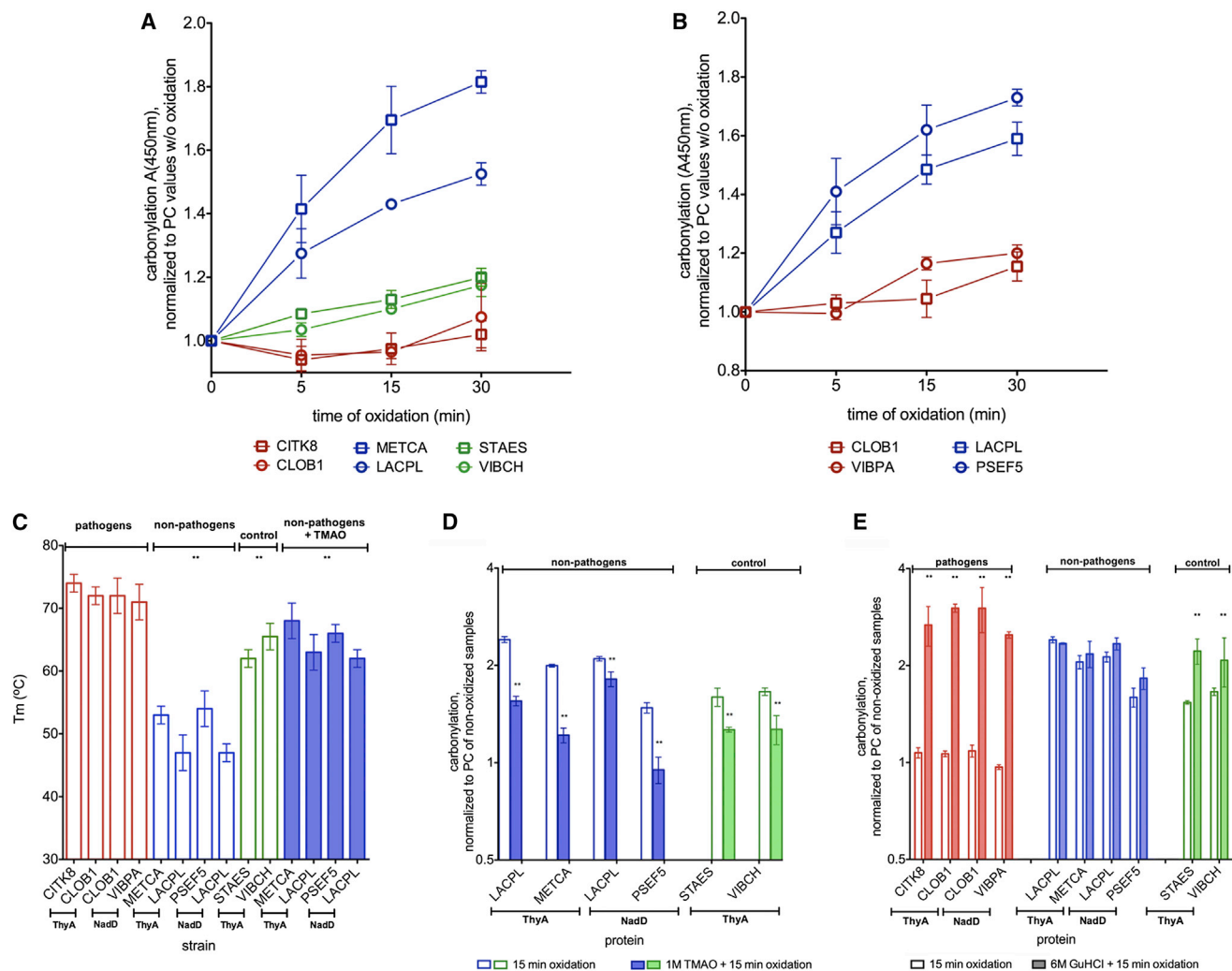


Figure 3. Oxidation Resistance and Conformational Stability of Proteins with Pathogen-like and Non-Pathogen-like PaPS

(A and B) Protein carbonylation was measured in (A) ThyA from CITK8, CLOB1, METCA, LACPL, STAES, and VIBCH and (B) NadD from CLOB1, VIBPA, LACPL, and PSEF5.

(C) Melting temperatures in proteins from (A) and (B).

(D and E) Oxidation resistance of (D) ThyA and NadD with non-pathogen-like and control PaPS upon incubation with 1 M TMAO prior to oxidation and (E) ThyA and NadD from all species listed above upon incubation with 6 M GuHCl prior to oxidation. Oxidation resistance is monitored as protein carbonylation oxidation, normalized to protein carbonylation values of nonoxidized proteins.

Error bars represent SD of three triplicate measurements. ** $p < 0.01$ (two-tailed t test).

Following Fenton-reaction-induced oxidation, the HTH (Figure 4A) and BH (Figure 4B) peptides with non-pathogen-like PaPS values exhibited an average 30% increase in carbonylation ($p < 0.0001$), unlike the peptides with pathogen-like PaPS, which did not exhibit an increase in their carbonylation level. The negative control peptides (which had a different overall composition while retaining non-pathogen-like PaPS values; see Experimental Procedures) again displayed oxidation resistance resembling the non-pathogen-like peptides. Furthermore, the peptides with pathogen-like PaPS displayed an increased conformational stability, with a T_m 13°C higher for HTH and 20°C higher for the BH than the corresponding non-pathogen-like peptides ($p < 0.0001$) (Figures 4C and S4D–S4F). The nega-

tive control peptides resembled the non-pathogen-like peptides; their T_m is 18°C higher than the pathogen-like peptides (Figures 4C and S4D–S4F).

Conformational stability of the synthetic peptides correlates with their oxidation resistance, thus establishing that the relationship between oxidation resistance and conformational stability is not limited to natural proteins.

DISCUSSION

Even though pathogenic bacteria are characterized by different mechanisms of how they cause disease, common hallmarks of infectivity exist across evolutionarily distant pathogenic species

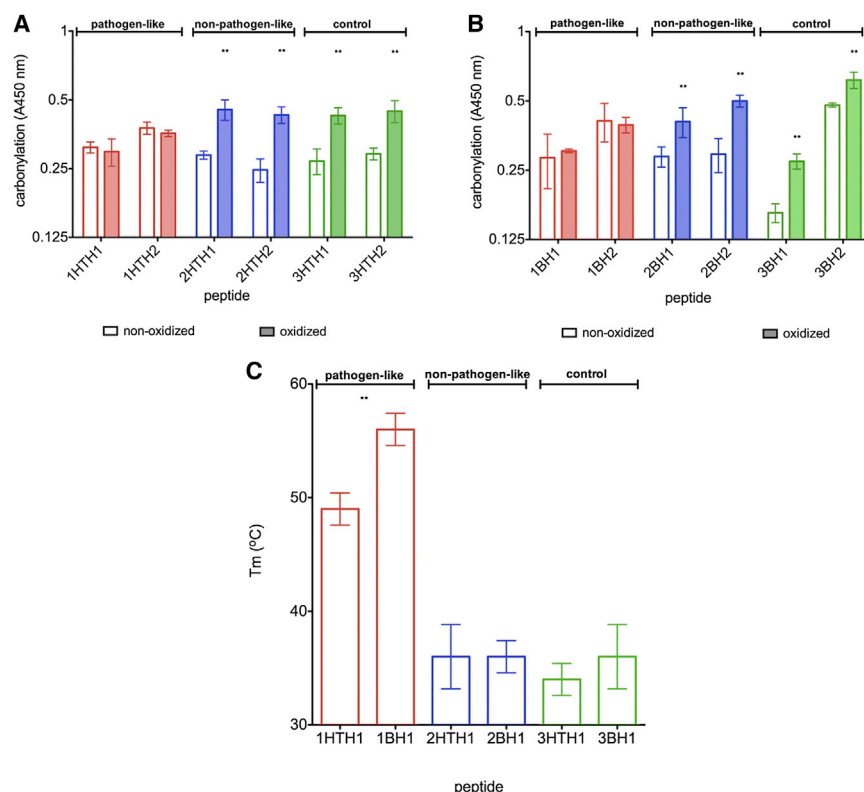


Figure 4. Oxidation Resistance of the Synthetic Pathogen-like and Non-Pathogen-like Peptides

(A and B) Oxidation resistance of (A) helix-turn-helix (HTH) and (B) β -hairpin (BH) designed peptides. Oxidation resistance is monitored as peptide carbonylation after 5 min of Fenton-induced oxidation.

(C) Melting temperatures of the synthetic peptides. Error bars represent the SD of three triplicate measurements. ** $p < 0.01$ (two-tailed t test).

(Gama et al., 2012). In our study, we investigated the characteristics of overall proteome composition shared between pathogenic bacteria and searched for a hypothesized link to tolerance to oxidative stress, a challenge faced by pathogens during host invasion (Rosenberger and Finlay, 2003).

As a consequence of oxidative damage, proteins are unable to maintain their native structure and thus perform their function with a decreased efficiency (Krisko and Radman, 2010), leading to elevated mutation rates and fueling a vicious cycle of protein and genetic damage (Krisko and Radman, 2013). Although proteins can undergo different oxidative modifications in a number of ways, carbonylation is a common irreversible type of damage elevated during oxidative stress (Nyström, 2005). Importantly, even in conditions of physiological ROS levels, misfolding of proteins increases their susceptibility to carbonylation in vivo, for instance during heat shock or streptomycin or puromycin treatments (Dukan et al., 2000). However, the natively folded proteins can also be carbonylated, and the individual proteins display extreme inequalities in their resistance to carbonylation, with $\sim 10\%$ of the *E. coli* proteome being highly susceptible (Maisonneuve et al., 2009). The sequence or structure determinants of these differences are not clear, but the strong tendency of carbonylated sites to cluster together indicates that some underlying principles must exist (Rao and Möller, 2011).

Our initial experiments have shown the increased resistance of the representative pathogenic bacteria to an HSP90 inhibitor and of their pure protein extracts to in vitro oxidation (see also Supplemental Discussion). These observations are consistent with the proteome properties summarized in the PaPS signature

that point toward an increased secondary structure content and a decreased proportion of disorder-promoting amino acids in pathogens. Highly ordered proteins contain a larger proportion of solvent-inaccessible amino acids involved in high-energy intramolecular interactions, and solvent accessibility has been discussed as a likely determinant of carbonylation-prone sites (Maisonneuve et al., 2009). Interestingly, we found pathogens to strongly disfavor proline, which not only disrupts secondary structures but also is known to be the most carbonylation-susceptible amino acid (Rao and Möller, 2011). However, the content of

oxidation-prone amino acids does not appear to consistently reflect the oxygen resistance of the organism (Vieira-Silva and Rocha, 2008). On the other hand, the increased proteome carbonylation in vivo under stresses leading to protein misfolding (Dukan et al., 2000) suggests that the proteins (or parts thereof) intrinsically more resistant to misfolding might also be more resistant to carbonylation. Finally, we demonstrate in *E. coli* that solely the level of ROS-scavenging enzymes or the presence of small-molecule scavengers (NAC) cannot explain the full extent of ROS resistance differences among pathogen proteomes (Figure S1). Rather, we propose that an intrinsic property of proteins is responsible for the extra ROS resistance not explained by other mechanisms we tested.

Our work provides multiple lines of experimental evidence that demonstrate a causal relationship between the conformational stability of a folded protein structure and its resilience to carbonylation. Furthermore, the PaPS score that estimates structural stability was derived from proteome composition of pathogenic bacteria, suggesting a link between PaPS and pathogenicity, although alternative (or complementary) explanations for PaPS are possible. Our data suggest that optimal growth temperatures, doubling times, or aerobicity does not fully explain PaPS (Figure S3). In addition, PaPS could be indirectly linked to pathogenicity via the genomic G+C content. Host-associated bacteria tend toward lower G+C than free-living ones, and thus pathogens also have somewhat lower G+C overall. Genomic G+C is known to influence the amino acid composition of the proteome (Moura et al., 2013), where low G+C corresponds to a proteome more stable against unfolding (Mendez et al.,

2010). Thus, the observed signature of conformational stability in pathogenic proteomes might be, in part, an indirect effect of selection (or relaxation thereof) on G+C content (reviewed in Rocha and Feil, 2010). Nevertheless, genomic G+C does not appear to fully explain our experimental observations; three of the six representative pathogens had moderately high (*S. typhimurium* [52%] and *K. pneumoniae* [57%]) or very high G+C (*P. aeruginosa* [67%]) but displayed similar geldanamycin resistance as the three lower G+C pathogens (Figure 1B) as well as similar proteome resistance to oxidation (Figures 1A and 1C). Pathogenicity is a multifaceted phenotype, and the G+C content and the associated proteome composition appear to be additions to the set of factors already known to influence pathogenicity.

To conclude, increased conformational stability may represent an additional mechanism by which proteins of pathogenic bacteria resist oxidation, providing an intrinsic resilience against the host defenses. Moreover, it is possible to design proteins with an enhanced conformational stability and, consequently, with an increased resistance to oxidative stress by applying the protein properties encoded in the PaPS statistic derived from pathogen proteomes. In addition to implications for synthetic biology, these findings could also contribute to the design of novel therapeutic approaches through destabilization of bacterial proteins. Such strategies could prove to be an effective means of delivering a synergistic attack on pathogen proteomes by combining protein destabilization with the ROS-inducing host defenses.

EXPERIMENTAL PROCEDURES

Wild-type *E. coli* MG1655 was cultured in liquid Luria-Bertani (LB) medium at 37°C (always supplemented with the appropriate antibiotic). Overnight cultures were diluted 100-fold in LB supplemented with isopropyl β -D-1-thiogalactopyranoside (0.5 mM) at 37°C to the exponential phase (optical density 600 = 0.2–0.3).

Genes encoding ThyA and NadD were synthesized and inserted into pJexpress401 vector after codon usage optimization for *E. coli* MG1655 (purchased from DNA 2.0). Histidine tag (6x) was fused at the protein C terminus. Table S1 contains the list of organisms from which ThyA and NadD originate as well as the list of species' full names from which total protein extracts were analyzed. PSEAE, STAAF, ECOD1, SALT1, and KLEP3 were grown in LB medium, PSEP1 in nutrient medium; LACC3 in MRS medium; SHEON in oxioid CM3 medium; and CLOAB, CLOPS, STRAG, and BACC1 on tryptic soy medium. All strains listed were grown at 37°C to 0.2 optical density at 600 nm. CLOAB and CLOPS were grown in an oxygen-free atmosphere.

For all strains, exponential cultures were exposed to 2 μ M geldanamycin (final) for 3 hr at 37°C. Serial dilutions were plated and survival determined after overnight incubation at 37°C. Aliquots were stored for measurement of protein carbonylation (Supplemental Experimental Procedures).

In order to measure the ROS production, we used two procedures: dihydro-rhodamine-123 was used for ROS quantification via measuring cell fluorescence with excitation at 500 nm and emission at 530 nm, and 2',7'-dichloro-fluorescein diacetate was used with subsequent analysis on FACSCalibur flow cytometer (Supplemental Experimental Procedures).

His-tagged ThyA and NadD were purified on nickel nitrilotriacetic acid Superflow columns (QIAGEN) according to the manufacturer's instructions. Purified proteins were concentrated by the Amicon Ultra 3K centrifugal filter devices (Millipore). Purified ThyA and NadD were diluted to 10 μ g/ml, loaded into Maxisorb wells (Nunc), and incubated overnight at 4°C to allow proteins to adsorb to the surface. Adsorbed proteins were oxidized for 5, 15, and 30 min. After oxidation, the protein carbonylation was measured (Supplemental Experimental Procedures).

The design of the synthetic peptides is described in Supplemental Experimental Procedures.

The computational analyses were performed on genomes from clades Firmicutes (22 nonpathogenic to vertebrates and 19 pathogenic; maximum of one strain per species) and γ -proteobacteria (18 and 27), where we controlled for varying content of gene families (details on genome and gene selection are given in Supplemental Experimental Procedures). The genes' translations to protein were described by amino acid relative frequencies and by 59 physicochemical properties (Smole et al., 2011) from the Amino Acid Index database (Kawashima and Kanehisa, 2000) and a PC analysis performed (using XLStat 2010.6, Addinsoft) across the unweighted averages of proteins for each proteome. PaPS was defined as a rotation of the first two PCs that maximizes the separation of the pathogen and nonpathogen proteomes and pathogen unrelated proteome signature (PUPS) as orthogonal to it; details are given in Supplemental Experimental Procedures and a visualization is provided in Figure S2. The ThyA and NadD gene families were selected among 853 widespread clusters of orthologous groups (COGs) (Tatusov et al., 2003) to maximize the area under the curve score for the separation of pathogens and nonpathogens by PaPS (but not PUPS) of proteins within that COG, consistently in Firmicutes and γ -proteobacteria (Supplemental Experimental Procedures).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Results, Supplemental Discussion, Supplemental Experimental Procedures, four figures, and four tables and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2014.04.057>.

ACKNOWLEDGMENTS

The authors thank Miroslav Radman for stimulating discussions and Tea Copic for technical support. This work was supported by the Mediterranean Institute for Life Sciences. A.V. and A.N. are supported by the City of Split Fellowship for Excellence. The work of F.S. was supported in part by Marie Curie Actions; by grants 00980501 and 098-0000000-3168 of the Ministry of Science, Education, and Sport of Croatia; and by EU grant ICT-2013-612944 (MAESTRA). A.K. would like to thank Maria Marta Radman for her help with the graphics.

Received: December 12, 2013

Revised: March 26, 2014

Accepted: April 30, 2014

Published: May 29, 2014

REFERENCES

- Dukan, S., Farewell, A., Ballesteros, M., Taddei, F., Radman, M., and Nyström, T. (2000). Protein oxidation in response to increased transcriptional or translational errors. *Proc. Natl. Acad. Sci. USA* 97, 5746–5749.
- Gama, J.A., Abby, S.S., Vieira-Silva, S., Dionisio, F., and Rocha, E.P. (2012). Immune subversion and quorum-sensing shape the variation in infectious dose among bacterial pathogens. *PLoS Pathog.* 8, e1002503.
- Imlay, J.A. (2003). Pathways of oxidative damage. *Annu. Rev. Microbiol.* 57, 395–418.
- Kawashima, S., and Kanehisa, M. (2000). AAindex: amino acid index database. *Nucleic Acids Res.* 28, 374.
- Krisko, A., and Radman, M. (2010). Protein damage and death by radiation in *Escherichia coli* and *Deinococcus radiodurans*. *Proc. Natl. Acad. Sci. USA* 107, 14373–14377.
- Krisko, A., and Radman, M. (2013). Phenotypic and genetic consequences of protein damage. *PLoS Genet.* 9, e1003810.
- Maisonneuve, E., Ducret, A., Khoeiry, P., Lignon, S., Longhi, S., Talla, E., and Dukan, S. (2009). Rules governing selective protein carbonylation. *PLoS ONE* 4, e7269.

- Mendez, R., Fritsche, M., Porto, M., and Bastolla, U. (2010). Mutation bias favors protein folding stability in the evolution of small populations. *PLoS Comput. Biol.* 6, e1000767.
- Moura, A., Savageau, M.A., and Alves, R. (2013). Relative amino acid composition signatures of organisms and environments. *PLoS ONE* 8, e77319.
- Nyström, T. (2005). Role of oxidative carbonylation in protein quality control and senescence. *EMBO J.* 24, 1311–1317.
- Rao, R.S., and Möller, I.M. (2011). Pattern of occurrence and occupancy of carbonylation sites in proteins. *Proteomics* 11, 4166–4173.
- Rocha, E.P.C., and Feil, E.J. (2010). Mutational patterns cannot explain genome composition: Are there any neutral sites in the genomes of bacteria? *PLoS Genet.* 6, e1001104.
- Rosenberger, C.M., and Finlay, B.B. (2003). Phagocyte sabotage: disruption of macrophage signalling by bacterial pathogens. *Nat. Rev. Mol. Cell Biol.* 4, 385–396.
- Smole, Z., Nikolic, N., Supek, F., Šmuc, T., Sbalzarini, I.F., and Krisko, A. (2011). Proteome sequence features carry signatures of the environmental niche of prokaryotes. *BMC Evol. Biol.* 11, 26.
- Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., et al. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4, 41.
- Tekaia, F., and Yeramian, E. (2006). Evolution of proteomes: fundamental signatures and global trends in amino acid compositions. *BMC Genomics* 7, 307.
- Uversky, V.N. (2013). Unusual biophysics of intrinsically disordered proteins. *Biochim. Biophys. Acta* 1834, 932–951.
- Vieira-Silva, S., and Rocha, E.P. (2008). An assessment of the impacts of molecular oxygen on the evolution of proteomes. *Mol. Biol. Evol.* 25, 1931–1942.
- Zeldovich, K.B., Berezovsky, I.N., and Shakhnovich, E.I. (2007). Protein and DNA sequence determinants of thermophilic adaptation. *PLoS Comput. Biol.* 3, e5.

Signatures of conformational stability and oxidation resistance in proteomes of pathogenic bacteria

Anita Vidovic¹, Fran Supek^{2,3,4}, Andrea Nikolic¹, Anita Krisko^{1,#}

¹ Mediterranean Institute for Life Sciences, Mestrovicova setaliste 45, 21000 Split, Croatia.

² Rudjer Boskovic Institute, Division of Electronics, Bijenicka cesta 54, 10000 Zagreb, Croatia.

³ EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain.

⁴ Universitat Pompeu Fabra (UPF), 08002 Barcelona, Spain.

Supplemental Experimental Procedures

Geldanamycin treatment

Exponential cultures of STAAF, SHEON, PSEAE, KLEP3, LACC3, PSEP1, ECOD1, CLOAB, CLOPS, STRAG, SALTI and BACC1 were exposed to 2 μ M geldanamycin (final) during 3h at 37°C. The same treatment was applied on *E. coli* MG1655 bearing pCA24N plasmids with *katE*, *katG* and *sodD* genes, separately. For induction, 100 μ M IPTG was used. Furthermore, *E. coli* was grown until exponential phase in the presence on 2.5 mM N-acetyl-cysteine (final) and exposed to geldanamycin as described above. Serial dilutions were plated and survival determined after overnight incubation at 37°C. Aliquots were stored for measurement of protein carbonylation.

To obtain protein extracts, the collected cells were pelleted by 15 minute centrifugation at 10,000 x g. Pellets were resuspended in the lysis buffer containing 10 mM PBS, and protease inhibitors (Roche). Preparation of extracts was followed by a 5 minute oxidation (25 mM sodium ascorbate and 0.1 mM FeCl₃ in 25 mM Hepes) and protein carbonylation measurement.

In order to measure the amount of geldanamycin that has entered the cells, geldanamycin fused with fluorescein isothiocyanate (FITC-GL) was used in the same conditions as described above for regular geldanamycin. Following the 3-hour incubation, the cells were washed in 10 mM PBS and their fluorescence was measured with excitation at 495 nm and emission at 519 nm.

Temperature increase resistance measurement

In order to experimentally examine the influence of moderate temperature shifts on live bacteria, exponential cultures of STAAF, SHEON, PSEAE, KLEP3, LACC3, PSEP1, ECOD1, CLOAB, CLOPS, STRAG, SALTI and BACC1 were exposed to 39°C, 41°C, 50°C and 60°C during 1 hour. Serial dilutions were plated and survival determined after overnight incubation at 37°C. Aliquots were stored for measurement of protein carbonylation.

ROS measurement

In order to measure the ROS production in all studied cells, bacteria were grown to OD 0.2-0.3 at 600 nm in the conditions described above. Geldanamycin was added (2 μ M final) and after 15 minute incubation at 37°C, dihydrorhodamine-123 was added (25 μ M final). After additional 30 minutes of incubation at 37°C, cells were washed in 10 mM PBS. The amount of produced ROS was evaluated by measuring cell fluorescence with excitation at 500 nm and emission at 530 nm.

For intracellular ROS analysis by flow cytometry, cells were incubated in the dark with 10 μ g/ μ L 2',7'-dichlorofluorescein diacetate (H2DCFDA, Sigma) for 120 min at 37°C and subsequently analyzed on FACSCalibur flow cytometer. The fluorescence of 10,000 cells resulting from the intracellular conversion of non-fluorescent H2DCFDA into fluorescent 2',7'-dichlorofluorescein (DCF) was measured in FL1 channel. The collected data was analyzed using FlowJo software version 7.2.5 for Microsoft (TreeStar, San Carlos, CA, USA) to determine the mean green fluorescence intensity after the each treatment. The results are expressed as % of mean fluorescence of the wild type control strain.

Protein extraction and oxidation in vitro

Bacterial cells were broken by using a mechanical homogenizer and glass beads (\leq 106 μ m). Cell lysates were centrifuged for 20 min at 10,000 x g. Samples were then supplemented with 10 mg/100 mL lipid removal agent (Sigma 13360-U), incubated for 1 hour at room temperature with shaking and pelleted for 15 min at 10,000 x g. The amount of protein in the supernatant was measured by the Micro

BCA™ Protein Assay Reagent Kit (Pierce). Protein extracts diluted to 10 µg/mL were loaded into Maxisorb wells (Nunc) and incubated overnight at 4°C to allow proteins to adsorb to the surface prior to their 5 minute oxidation and protein carbonylation measurement (Supplemental Experimental Procedures). Adsorbed proteins were oxidized for 5 minutes using the oxidation solution containing 25 mM Hepes buffer, 25 mM sodium ascorbate, and 0.1 mM FeCl₃.

Protein Carbonylation Measurement

Adsorbed proteins were derivatized by using 12 µg/mL DNPH. Derivatization of adsorbed proteins is followed by detection of derivatized dinitrophenol (DNP)-carbonyl by a rabbit anti-DNP primary antibody (Sigma, D9656) and goat anti-rabbit secondary antibody conjugated to HRP (Jackson ImmunoResearch, 111-035-14). Stocks of antibodies were prepared at ~1 µg/µL and used at 1:7000 dilutions. Subsequent incubation with enzyme substrate 3,3',5,5'-tetramethylbenzidine resulted in a colored product that was quantified using a microplate reader with maximum absorbance at 450 nm.

Purification of ThyA and NadD and treatment with GuHCl and TMAO

Exponential *E. coli* cultures bearing plasmids with ThyA or NadD were collected and treated as described above. The lysis buffer was supplemented with 20 mM imidazole and lysozyme (1 mg/mL). Samples were incubated for 30 min at 37°C and centrifuged for 30 min at 10,000 x g. To remove large proteins from the sample, the Amicon Ultra 50K centrifugal filter devices (Millipore) were used (filtrate was collected). Ni-NTA Superflow columns (Qiagen) were used for purification of 6xHis-tagged ThyA and NadD according to manufacturer's instructions. Purified proteins were concentrated by the Amicon Ultra 3K centrifugal filter devices (Millipore). Purified ThyA and NadD were diluted to 10 µg/mL, loaded into Maxisorb wells (Nunc) and incubated overnight at 4°C to allow proteins to adsorb to the surface. Adsorbed proteins were oxidized for 5, 15 and 30 minutes. After oxidation, the protein carbonylation was measured.

In addition, all purified ThyA and NadD were treated with 6M guanidinium hydrochloride (GuHCl) for 5 minutes at room temperature. ThyA from LACPL, STAES, METCA and VIBCH (nonpathogens and controls), and NadD from LACPL and PSEF5 (non-pathogens) were treated with 1M trimethylamine N-oxide (TMAO) for 1.5 hour at 37°C. These treatments preceded protein adsorption onto Maxisorb wells, 15 minute oxidation and carbonylation measurement.

Purity of ThyA or NadD was confirmed by SDS-PAGE (12% (v/v) polyacrylamide) with Coomassie staining. A band corresponding to 35 kDa for ThyA protein and 23 kDa for NadD protein was found in the gel.

Thermal denaturation

Purified ThyA and NadD as well as the synthetic peptides were exposed to thermal denaturation by the increasing temperature from 30°C to 94°C in 2°C steps. At each temperature, absorbance at 280 nm was measured to obtain a protein melting curve for each protein. Conformational stability of proteins was inferred from this data by estimating of the melting temperature (T_m) at which 50% of the protein is denatured. Importantly, the *E. coli* extract is a mixture of proteins that are characterized by a wide variety of conformational stabilities, ranging from unstable to very stable ones, thus the melting curve stretches across the entire temperature range.

Design of synthetic peptides

Two series of peptides were designed with high propensity to fold into (i) helix-turn-helix (HTH), and (ii) beta-hairpin (BH) motif; the design of new peptides was based on parts of previously described

HTH and BH containing proteins. Average length of all peptides was ~25 aa. The exact amino-acid sequence of each peptide was designed based on PaPS and PUPS values as follows: (1) peptides with PaPS values characteristic for proteins from pathogenic bacteria, (2) peptides with PaPS values characteristic for proteins from non-pathogenic bacteria, and (3) peptides with again non-pathogen-like PaPS values, but with PUPS values different from the proteins in group (2), thus serving as a control that a change in amino acid composition that is unrelated to PaPS does not affect protein conformational stability or oxidation resistance. Peptides were synthesized and purchased from GeneCust. Names of peptides together with their sequence properties are summarized in Table S3. For all peptides we have measured their oxidation resistance and conformational stability as described above.

In all cases, unless explicitly noted differently, the statistical significance of the obtained results has been determined by using two-tailed t-tests. In the name of the peptide, the number in front of the letters denotes (A) a peptide with pathogen-like PaPS, (B) a peptide with non-pathogen-like PaPS, and (C) control peptide. The peptides were designed by using pre-existing peptides already known to be alpha-helices and beta-hairpins as building blocks.

Quantifying principal trends in the proteomes' physicochemical properties.

From the fully sequenced bacterial genomes in the NCBI Genome Projects [<ftp://ftp.ncbi.nih.gov/genomes/Bacteria/>], we selected bacteria (a) clearly annotated as either pathogenic to vertebrates, or non-pathogenic to vertebrates, and (b) belonging to either of the two large phylogenetic groups: Firmicutes (22 non-pathogenic and 19 pathogenic) and γ -proteobacteria (18 and 27). Only a single strain was allowed for each bacterial species to avoid biasing the analysis toward heavily re-sequenced microbes. Reduced genomes of intracellular parasites or endosymbionts were excluded due to effects of extreme GC content on amino acid composition.

The pathogen versus non-pathogen annotation was taken from the NCBI Genome Projects "lproks0" table [ftp://ftp.ncbi.nih.gov/genomes/genomeprj/lproks_0.txt] followed by manual curation. Opportunistic pathogens were considered as pathogenic, e.g. *C. koseri*.

Before further comparison of pathogen to the non-pathogen proteomes, we performed a control for genomic content of gene families. Larger bacterial genomes are enriched with certain gene functional categories (e.g. transcription factors) and smaller ones are enriched in indispensable core functions (e.g. translation) (Ranea, et al., 2005; Supek, et al., 2010). This might confound an analysis of proteome physicochemical properties as some functional categories might have a peculiar amino acid composition, for instance the ribosomal proteins (Lin, et al., 2002). To control for such influences, first, we keep only the genes that could be assigned to either a COG (Tatusov, et al., 2003) or a NOG orthologous group from the eggNOG database (Jensen, et al., 2008), amounting to an average of 2986.7 (of original 3430.3) genes per genome. Second, within the Firmicutes and the γ -proteobacteria separately, we sampled the genes to enforce an equal frequency of each COG/NOG in the pathogens and in the non-pathogens, retaining an average 2352.5 genes per genome.

Next, the genes' translations to protein were described by 59 physicochemical properties used in a previous study of extremophile proteins (Smole, et al., 2011), mostly based on a previous selection (Atchley, et al., 2005) from the Amino Acid Index database. Proteomes were represented by an unweighted average of the values of all sampled proteins. A principal components analysis was performed using XLStat 2010.6 on the 59 properties and the 20 amino acid frequencies - total 79 features - across the 86 proteomes. The "Pathogen Proteome Signature" (PaPS) is a linear combination of PC1 and PC2 that has a slope (0.66) determined as an average of the slopes of the lines connecting the pathogen and non-pathogen centroids in Firmicutes (0.72) and in γ proteobacteria (0.59). Thus, PaPS is the direction of variability in the physicochemical property space that maximizes the separation of the

pathogen and non-pathogen proteomes. For finding the physicochemical properties correlated to PaPS (Figure 2D), a nonredundant set was used (<0.9 Spearman correlation between any pair of properties in set); details in Additional file 6 in (Smole, et al., 2011).

Additionally, we also designed a control variable PUPS (“pathogen-unrelated proteome signature”) that is uncorrelated to the PaPS on the principal components plots: it has a slope equal to the negative reciprocal slope of the PaPS in this coordinate system (0.66 for PaPS; -0.34 for PUPS; effectively, the PaPS and the PUPS are rotations of the original PC1 and PC2 axes, respectively). Thus, two proteomes with different PUPS values are guaranteed to have a different amino acid composition in a way that does not affect their pathogen-likeness if they have equal PaPS values.

Selecting gene families that display strong pathogen vs. non-pathogen differences

In order to proceed with experimental validation, we selected specific gene families that faithfully represent the whole-proteome trends between pathogens and non-pathogens summarized in PaPS. We thus evaluated 853 widespread COGs (those having at least 10 representative proteins in each of the following: pathogenic Firmicutes, pathogenic γ -proteobacteria, non-pathogenic Firmicutes and non-pathogenic γ -proteobacteria). To determine how well PaPS separates the pathogen from the non-pathogen proteomes within each COG, we calculated area-under-ROC-curve (AUC) scores for COGs, separately for the two phylogenetic groups. The AUC scores were also independently calculated for the PUPS control signature. An AUC of ~ 0.5 corresponds to no separation, and of 1.0 to perfect separation. To experimentally test our hypothesis that a difference in the PaPS signature, but not the PUPS control signature, confers conformational stability and oxidation resistance to the pathogens’ proteins, we selected two COGs: COG0207 (ThyA in *E. coli*) and COG1057 (NadD in *E. coli*) with a high AUC score for PaPS, but a low AUC score for PUPS, in both phylogenetic groups. In particular, ThyA had $AUC_{PaPS}=0.80$ and 0.83 , and $AUC_{PUPS}=0.50$ and 0.48 , for Firmicutes and γ -proteobacteria, respectively. For NadD, the $AUC_{PaPS}=0.87$ and 0.79 and $AUC_{PUPS}=0.60$ and 0.43 . These AUC scores correspond to a statistically significant separation in PaPS between the pathogen and non-pathogen proteins (P-values by Mann-Whitney U test in Figures 2E-H). The two COGs were chosen to also satisfy other criteria: ease of heterologous expression, sufficient protein length, as well as length homogeneity within the COG family.

For each of the ThyA and NadD gene families, we proceeded to select representative proteins based on the proteins’ PaPS values. We selected ThyA proteins: (a) two from pathogens (CLOB1, CITK8) with pathogen-like PaPS values; (b) two from non-pathogens (LACPL, METCA) with nonpathogen-like PaPS; and (c) two from pathogens (STAES, VIBCH) with nonpathogen-like PaPS similar to the last group, but here with different overall amino acid composition (the PUPS control score). The control group (c) verifies that it is specifically the PaPS signature in the amino acid composition that determines the oxidation resistance. Furthermore, since these proteins come from pathogens but have non-pathogen-like PaPS, they also verify that the PaPS determines the pathogen-likeness of the protein, and not some other property of the pathogen proteomes. For NadD, we selected: (a) two proteins from pathogens (CLOB1, SHEON) with pathogen-like PaPS and (b) two from non-pathogens (LACPL, PSEF5) with nonpathogen-like PaPS. To ensure that the observed effects are not phylogenetic group-specific, these pairs of species always consist of one Firmicutes and one γ -proteobacterium.

Supplemental Results

Geldanamycin penetration and ROS measurements

The observed distinction in geldanamycin resistance between the representative pathogens and nonpathogens may have resulted from an unequal geldanamycin membrane permeability among the studied species. Therefore, we measured the intracellular level of geldanamycin using a fluorescent derivative, fluorescein isothiocyanate-geldanamycin (Figure S1A). The fluorescence signal indicated that approximately the same amount of geldanamycin entered into the cells of all studied bacterial species, ruling out a differential membrane permeability as a cause for the low susceptibility of pathogens to geldanamycin.

In addition, geldanamycin-induced mortality might originate from ROS production rather than the protein folding effects. To investigate this aspect, we exposed bacteria to geldanamycin in the presence of intracellular DHR-123, a ROS-sensitive fluorescent dye. The presence of geldanamycin did not cause increase in ROS production in any of the tested species as observed both by fluorescence measurements in solution (Figure S1B) and by flow cytometry (Figure S1C). Thus, the possibility that the pathogens' resistance to geldanamycin relies on efficient suppression of geldanamycin-induced ROS is excluded.

Protective effect of ROS-scavenging activities towards geldanamycin

Herein, we set out to experimentally evaluate the protective effect to geldanamycin exposure granted by the antioxidant enzymes and by small molecule antioxidant N-acetylcysteine in *E. coli*, in comparison to the increase in geldanamycin resistance we had previously observed between the pathogenic bacteria relative to the non-pathogens. Overexpressing KatE and KatG in *E. coli* granted its proteome only 28 % and 15 % of the protection against geldanamycin (Figure S1D), expressed as a percentage of the average difference between the pathogenic and non-pathogenic species (Figure 1). In particular, the KatE overexpressing *E. coli* exhibited 0.81x of the protein carbonylation and 1.85x the geldanamycin survival of the wild type *E. coli*; the KatG overexpressing *E. coli* exhibited 0.87x of the protein carbonylation and 1.45x the geldanamycin survival (Figure S1E). To verify that the observed data are not specific to catalase, we also overexpressed superoxide dismutase SodA in *E. coli*, yielding 18 % of the protection against geldanamycin, again normalized to the difference between pathogenic and non-pathogenic species (Figure S1D). SodA overexpressing *E. coli* exhibited 0.93x of the protein carbonylation and 1.55x the geldanamycin survival of the wild type *E. coli* (Figure S1E).

Finally, in order to check also if the presence of an antioxidant would result in a similar effect as observed between pathogens and non-pathogens, we have pre-treated *E. coli* with 2.5 mM NAC and exposed it to geldanamycin, yielding only 10 % of the protection against geldanamycin relative to the one observed between pathogenic and non-pathogenic species (Figure S1D). *E. coli* pre-treated with NAC exhibits 0.89x of the protein carbonylation and 1.3x the geldanamycin survival of the wild type *E. coli* (Figure S1E). For comparison, the average differences between the 6 pathogenic and 6 non-pathogenic bacteria are 0.52x and 3.01x, in the carbonylation and geldanamycin survival assay, respectively.

Resistance to temperature increase is not related to the species' pathogenicity

In addition to the pathogenicity of the organisms, other explanations for the observed trends in proteome compositions are possible. Most prominently, one would expect mammalian (and avian) pathogens to be adapted to higher growth temperatures - ones that match their hosts' bodily temperatures, which are typically above the ambient temperatures. Moreover, one would expect the pathogens to be adapted to the temperature shifts resulting from the host responses to infection. We thus

statistically evaluate if the thermal preference of organisms can better explain the PaPS proteome score than the pathogenicity status can explain the PaPS. To this end, we employ ANCOVA, a regression technique that can simultaneously use both continuous variables (such as growth temperatures) and categorical variables (such as pathogenicity) to predict a ‘target variable’

(here, the PaPS proteome composition score). The coefficients (‘slopes’) on the explanatory variables (growth temperature, pathogenicity, etc.) can then be compared or tested for significance. We apply this test separately to 41 Firmicutes and the 45 Gamma-proteobacteria to examine if the pathogenicity, optimal growth temperature, range of temperatures permissive for growth, and several other lifestyle features (Figure S3A-D) predict the PaPS signature consistently across the two different evolutionary lineages.

In Firmicutes, we found that the pathogenicity of an organism can predict the PaPS proteome signature independently of the other examined features (standardized regression coefficient $b=0.46$, 95% CI: 0.18-0.75) with high significance ($P=0.002$). On the contrary, the optimal growth temperature could not predict PaPS ($b=-0.06$, $P>0.2$); also, the spread between the high and low points in the growth temperature range, an indicator of the ability of an organism to adapt to temperature shifts, could not predict PaPS ($b=0.02$, $P>0.2$). Next, we performed the same test with pathogenicity and the six other lifestyles also for the Gamma-proteobacteria, where we find a qualitatively similar result: pathogenicity is the strongest predictor of the PaPS proteome signature ($b=0.31$, $P=0.09$; next best is aerobicity, $b=0.19$), and pathogenicity is the only predictor significant at $P<0.1$ (all others, including growth temperatures have $P>0.2$). The standardized regression coefficients b for these lifestyle indicators, their 95% C.I. and the significance calls are given as Figure S3A, S3B.

Thermophiles are rare in our examined set of organisms - only 7/81 species, of that 2/7 pathogenic (this definition includes also the moderate ‘mesothermophiles’ with optimal growth temperatures $>40^{\circ}\text{C}$; see Table S2 for a list of organisms and phenotypes). Even so, it is still possible that the pathogens could have evolved a thermophile-like proteomic signature due to other kinds of environmental insults (e.g. ROS) that would challenge the proteome in a similar way as the high temperature does. To examine this possibility, we directly compared the PaPS signature to a known signature of thermophilic proteomes: thermophiles can be accurately distinguished by the sum of the frequencies of IVYWREL amino acids (Zeldovich, Berezovsky & Shakhovich, PLOS Comp Biol 2007).

- First, we checked whether the IVYWREL amino acids tend to have a substantially different contribution to the PaPS score than the other 13 amino acids. The difference between the median PaPS regression coefficients (from Figure 2C) of the IVYWREL set (-0.018 ± 0.135 , median \pm IQR) *versus* the remaining amino acids (0.020 ± 0.156) is quite small compared to the spread (distributions in Figure S3C). Thus, the PaPS and IVYWREL signatures don’t appear to be robustly linked across many amino acids.
- Second, we checked whether the pathogens differ from the non-pathogens in IVYWREL content of their proteomes, and we find no substantial difference (pathogens: $40.0\pm0.9\%$; non-pathogens $39.3\pm1.5\%$ [median \pm IQR] Figure S3D).

In addition to the analysis above, we turn to experimentally evaluate whether the 6 representative pathogenic bacteria employed in this study also possessed properties that are typically ascribed to thermophilic organisms.

To this end, we experimentally examined the influence of moderate temperature shifts on live bacteria, namely, the survival of the 6 representative pathogens and 6 non-pathogens at 39°C , 41°C , 50°C and 60°C . We have observed that all species, regardless of their pathogenicity, readily withstand the temperature increase to 39°C and 41°C , with average survival of 94% and 80%, respectively (Figure

S3E). On the other hand, larger temperature increases have significantly compromised bacterial survival: at 50°C and 60°C average survival of 18% and 0.07%, respectively, has been observed (Figure S3F). Although differences in survival can be observed between the individual studied species, they cannot be attributed to the pathogenicity of the species. Notably, the two anaerobic species (*Clostridium acetobutylicum* and *Clostridium perfringens*) do not display a different behaviour from the rest of the representative species.

In order to verify that the cell mortality observed at 50°C and 60°C correlates with the amount of protein damage, we measured total protein carbonylation levels in 12 representative species after the exposure to the increased temperatures. It can be seen (Figure S3F) that the increase in mortality is followed by an increase in the protein carbonylation: at 50°C and 60°C the average level of carbonylation is 2.6-fold and 3.6-fold higher than at 37°C, respectively. However, there are no observable differences between the species that could be explained by the species' pathogenicity.

Supplemental Discussion

Bactericidal effects of some antibiotics may be, at least in part, mediated by oxidative stress (Kohanski, et al., 2007). To examine this possibility, we measured the ROS levels in each representative species, in normal conditions and in the presence of geldanamycin. Regardless of the pathogenicity of the species and of its oxygen tolerance, ROS levels in the presence of geldanamycin appeared unchanged, including two obligate anaerobes, *C. perfringens* (pathogen) and *C. acetobutylicum* (non-pathogen). When considered together with the ROS levels, the higher geldanamycin-induced mortality in *C. acetobutylicum* relative to *C. perfringens* argues against its specificity to aerobes. In summary, these results exclude the possibility that the pathogens' geldanamycin resistance is due to efficient ROS suppression, but is most likely a consequence of an increased resistance of the pathogens' proteomes to HSP90 inhibition. Indeed, a relationship between an increased protein conformational stability and their resistance to oxidation in the selected bacteria is consistent with previous research showing misfolded proteins to be highly sensitive to oxidation *in vitro* (Dukan, et al., 2000).

Notable exceptions from the differences in proteome composition of pathogens

On overall, the PaPS scores were highly significantly different between the pathogens and nonpathogens (Figures 2A and 2B). There were, however, also exceptions to this trend. Such cases may reflect the subtleties in the definition of pathogenicity which are hard to capture using a binary classification, such as the one we applied. In particular, *Bacillus licheniformis* is a pathogen with the least pathogen-like PaPS in its group, classifying it with the non-pathogens (Figure S2A, S2B), and similarly so the *Pseudomonas mendocina* (second least pathogen-like PaPS, Figure S2A, S2B). In both cases, these are very lowly virulent organisms that can cause diseases in humans, but are known to do so extremely rarely (Johansen, et al., 2011; www.epa.gov/biotech_rule/pubs/fra/fra005.htm). In the same vein, we defined a pathogen as a microbe that infects vertebrates. Thus, two bacteria in our set that are known to commonly target insects (Waterfield, et al., 2004) were labeled as non-pathogens: *Bacillus thuringiensis* and *Photobacterium luminescens*. In both cases, however, the PaPS of their proteomes placed them in the pathogen range (Figure S2A, S2B). This suggests that the changes in proteome composition are consistent among a broader set of hosts than encompassed by our definition.

Furthermore, an interesting exception is *Listeria innocua*, a non-pathogenic organism having a pathogen-like proteome, and thus clustering together with its pathogenic relative, *Listeria monocytogenes* (highlighted on Figure S2A, S2B). Strikingly, *L. innocua* is known to have diverged from *L. monocytogenes* very recently (~29 MYa) by the loss of several pathogenicity genes (Hain, et al., 2006; den Bakker, et al., 2010). Apparently, it has still retained the pathogenic proteome composition, but has lost its ability to cause disease as it lacked specific genes that enable it to do so. A second non-pathogenic *Listeria* genome, *L. welshimeri*, followed the same evolutionary history (Hain, et al., 2006; den Bakker, et al., 2010) and also clusters together with *L. monocytogenes* by proteome composition (Figure S2). These three *Listeria* illustrate how a combination of genome-encoded features may be necessary to explain the pathogenic phenotype, and the proteome composition is one of such features - whose importance is underscored by its apparent prevalence among the examined microbes, exceptions notwithstanding.

Supplemental References

- Atchley, W.R., Zhao, J., Fernandes, A.D., Drüke, T. (2005). Solving the protein sequence metric problem. *Proc. Natl. Acad. Sci. U S A* 102, 6395-6400.
- den Bakker, H.C., Cummings, C.A., Ferreira, V., Vatta, P., Orsi, R.H., Degoricija, L., Barker, M., Petrauskene, O., Furtado M.R. and Wiedmann M. (2010). Comparative genomics of the bacterial genus *Listeria*: Genome evolution is characterized by limited gene acquisition and limited gene loss. *BMC Genomics*. 11, 688.
- Dukan, S., Farewell, A., Ballesteros, M., Taddei, F., Radman, M. and Nystrom, T. (2000). Protein oxidation in response to increased transcriptional or translational errors. *Proc. Natl. Acad. Sci. USA* 97, 5746-5749.
- Hain, T., Steinweg, C., Kuenne, C.T., Billion, A., Ghai, R., Chatterjee, S.S., Domann, E., Kärst, U., Goesmann, A., Bekel, T., Bartels, D., Kaiser, O., Meyer, F., Pühler, A., Weisshaar, B., Wehland, J., Liang, C., Dandekar, T., Lampidis, R., Kreft, J., Goebel, W. and Chakraborty, T. (2006). Wholegenome sequence of *Listeria welshimeri* reveals common steps in genome reduction with *Listeria innocua* as compared to *Listeria monocytogenes*. *J. Bacteriol.* 188, 7405-7415.
- Jensen, L.J., Julien, P., Kuhn, M., von Mering, C. and Muller, J. (2008). eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res.* 36, 250-254.
- Johansen, H. K., Kjeldsen, K., Høiby, N. (2011). *Pseudomonas mendocina* as a cause of chronic infective endocarditis in a patient with situs inversus. *Clin. Microbiol. Inf.* 7, 650-652.
- Kawashima, S., and Kanehisa, M. (2000). AAindex: amino acid index database. *Nucleic Acids Res.* 28, 374.
- Kohanski, M.A., Dwyer, D.J., Hayete, B., Lawrence, C.A. and Collins, J.J. (2007). A common mechanism of cellular death induced by bactericidal antibiotics. *Cell* 130, 797-810.
- Lin, K., Kuang, Y., Joseph, J.S., Kolatkar, P.R. (2002). Conserved codon composition of ribosomal protein coding genes in *Escherichia coli*, *Mycobacterium tuberculosis* and *Saccharomyces cerevisiae*: lessons from supervised machine learning in functional genomics. *Nucleic Acids Res* 30, 2599-2607.
- Ranea, J.A., Grant, A., Thornton, J.M. and Orengo, C.A. (2005). Microeconomic principles explain an optimal genome size in bacteria. *Trends Genet* 21, 21-25.
- Smole, Z., Nikolic, N., Supek, F., Šmuc, T., Sbalzarini, I.F. and Krisko A. (2011). Proteome sequence features carry signatures of the environmental niche of prokaryotes. *BMC Evol. Biol.* 11, 26.
- Supek, F., Skunca, N., Repar, J., Vlahovicek, K., Smuc, T. (2010). Translational selection is ubiquitous in prokaryotes. *PLoS Genet* 6, e1001004.
- Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., Smirnov, S., Sverdlov, A.V., Vasudevan, S., Wolf, Y.I., Yin, J.J., Natale, D.A. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4, 41.
- Waterfield, N.R., Wren, B.W. and Ffrench-Constant, R.H. (2004). Invertebrates as a source of emerging human pathogens. *Nat. Rev. Microbiol.* 2, 833-841.

Supplemental Figures and Legends

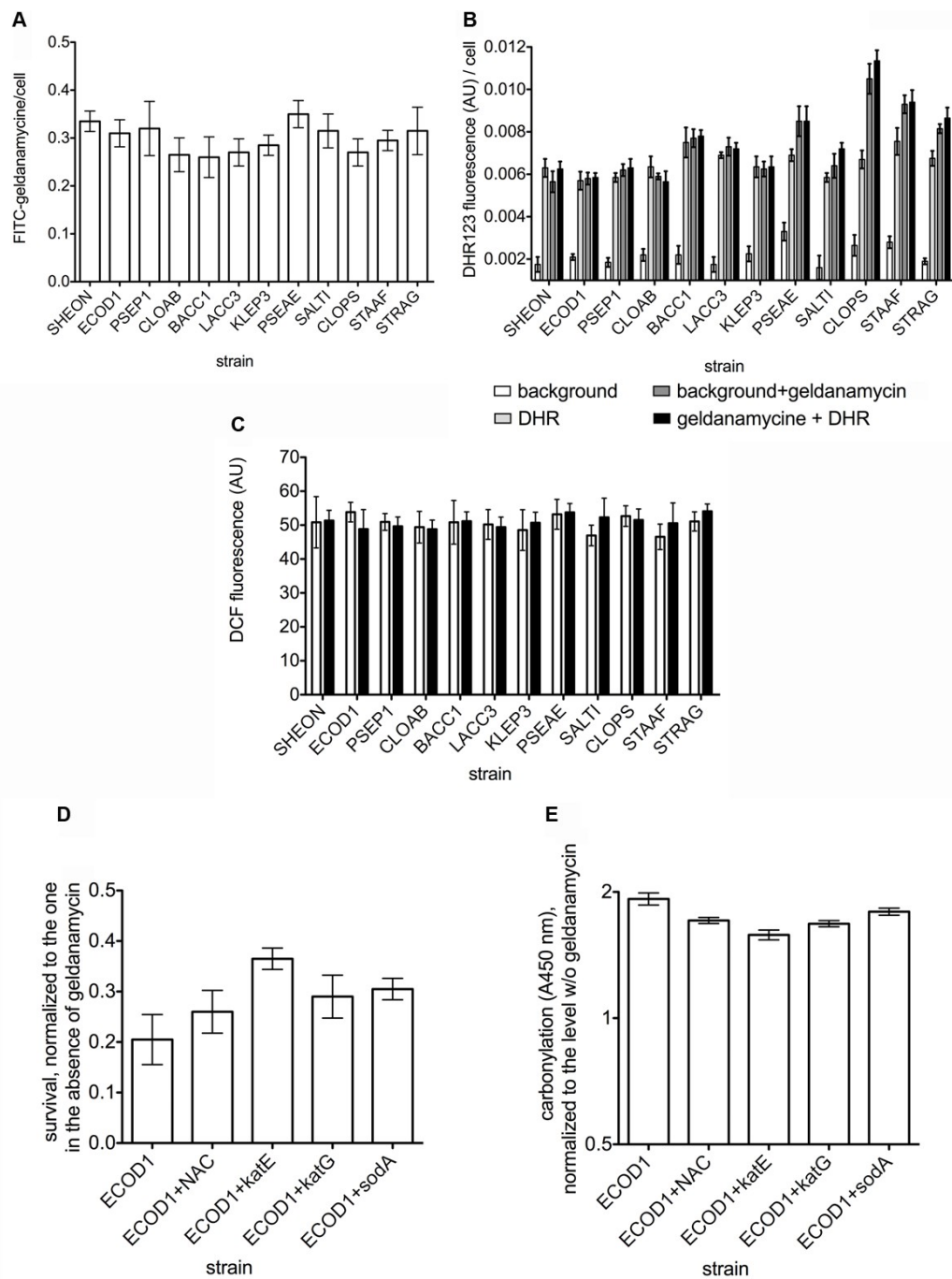


Figure S1, related to Figure 1. Fluorescence level (A) in the presence of FITC-labeled geldanamycin, **(B)** DHR 123 fluorescence in the absence and presence of geldanamycin, and **(C)** DCF fluorescence in the absence and presence of geldanamycin in SHEON, ECOD1, PSEP1, CLOAB, BACC1, LACC3, KLEP3, PSEAE, SALT1, CLOPS, STAAF and STRAG. **(D,E)** HSP90 inhibition in *E. coli* with enhanced antioxidative activity. **(D)** survival, and **(E)** total protein carbonylation in the presence of 2 μ M geldanamycin of *E. coli* with overexpressed Kat E, KatG, SodD and pretreated with N-acetylcysteine. Error bars represent the standard deviation of three triplicate measurements. ** denotes $p < 0.01$ (t-test, two-tailed).

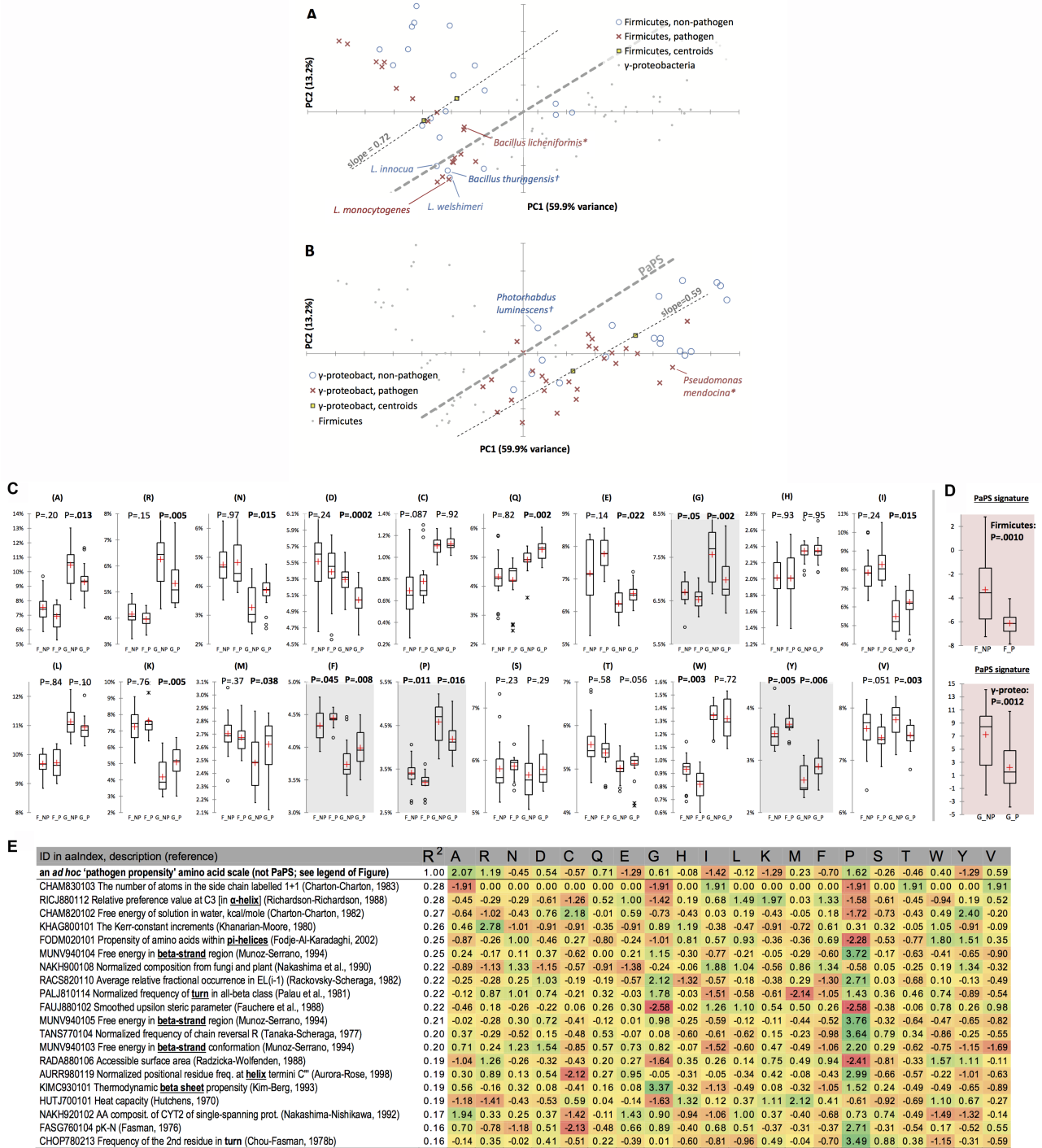


Figure S2, related to Figure 2. A principal components plot of the physicochemical properties of 86 proteomes. Visual emphasis is placed on either Firmicutes (A) or γ-proteobacteria (B). Dashed lines connect the pathogen and the non-pathogen centroids in each clade. The PaPS signature is a rotation of the coordinate system defined by the two PCs, where the first rotated PC (≡PaPS) has the average slope of the two centroid-connecting lines. Pathogen proteomes tend toward lower PaPS. Very weakly pathogenic species are marked with asterisks (*). Bacteria pathogenic only to insects are marked with daggers (†). “L.” stands for *Listeria*. See Supplemental

Material for a discussion of the highlighted species. **(C-E)** Usage of individual amino acids compared between pathogen and non-pathogens. **(C)** Distributions of amino acid relative frequencies, visualized separately in proteomes of Firmicutes (F_NP and F_P denote non-pathogens and pathogens, respectively) and γ -proteobacteria (G_NP and G_P). Boxes show median and 1st-3rd quartile, red crosses are averages, whiskers are minima and maxima after excluding outliers, which are shown individually as circles (or stars, in case of extreme outliers). P values are significance calls for differences between pathogens and non-pathogens, by two-tailed Mann-Whitney test. The amino acids where both Firmicutes and γ -proteobacteria are significant at $P \leq 0.05$ (corresponding to a joint $P \leq 0.017$ or a $FDR \leq 8.7\%$) are highlighted. **(D)** Distributions of PaPS scores for Firmicutes (above) and γ -proteobacteria (below); box plots and significance tests as in (C). **(E)** An *ad hoc* 'amino acid pathogen propensity' scale was defined by taking, for each amino acid, the \log_2 ratio of its average frequency in non-pathogens *vs.* its average frequency in pathogens, separately in Firmicutes and γ -proteobacteria. Then, the smallest absolute value of the two phyla was chosen for each amino acid to ensure the scale is supported across phylogenetically diverse organisms. This scale (first row of numbers) was compared by squared Pearson correlation (R^2) to all 484 amino acid scales in the *aaIndex* database and 20 highest correlating scales shown here (other rows). All scales are standardized across amino acids to mean=0 and standard deviation=1. Nine of 20 scales are related to protein structural elements (highlighted in description of scale).

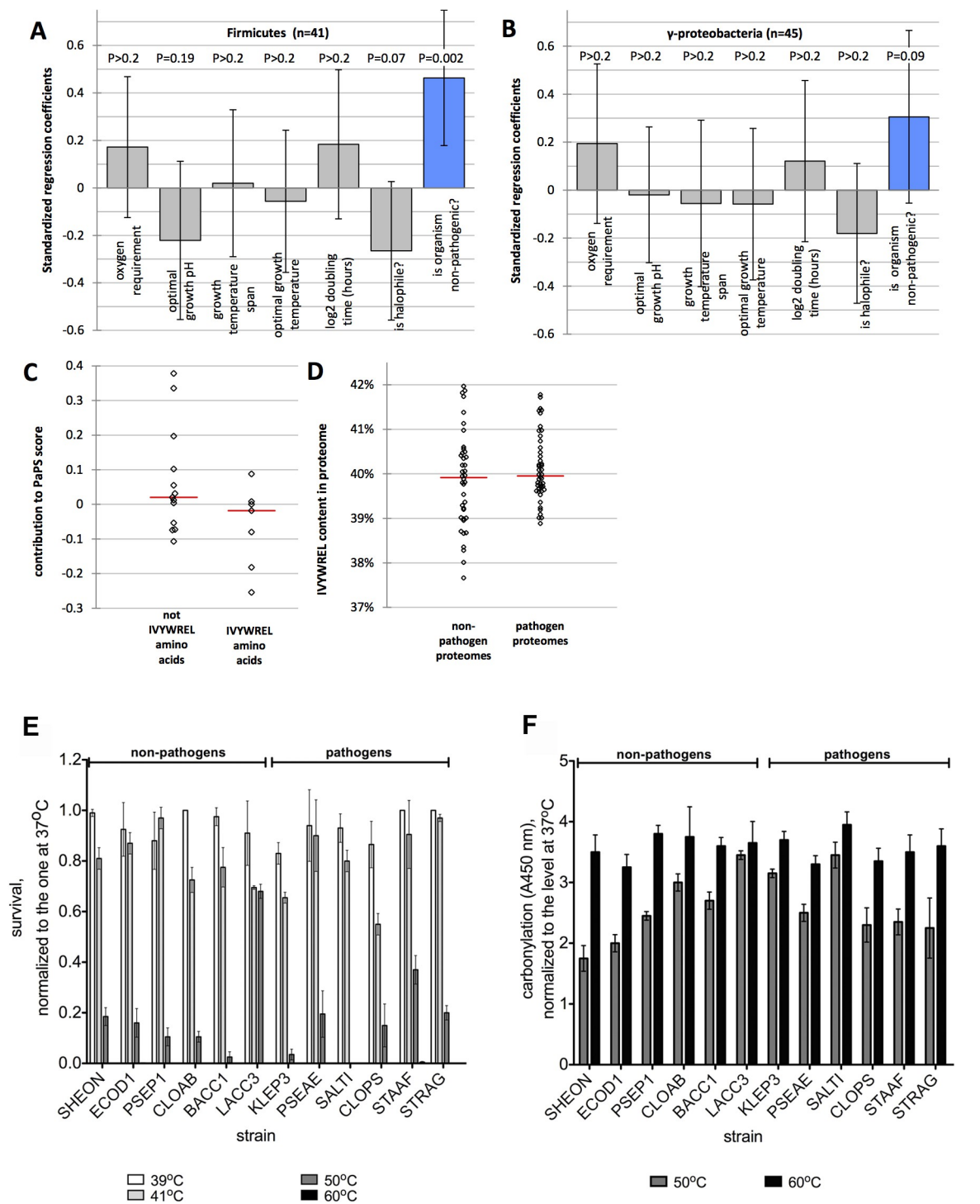


Figure S3, related to Figures 1 and 2. Checks for influence of confounding phenotypes on the PaPS proteome signature. (A, B) The standardized regression coefficients when regressing the PaPS levels of proteomes on seven phenotypic features of the bacteria, including the pathogenicity status and six possible confounders: four continuous variables (optimal pH, optimal growth temperature, growth temperature span [high minus low] and log₂ doubling time) and two categorical variables (halophilicity encoded as 0/1, and oxygen requirement encoded as: anaerobe or microaerophilic = 0, aerotolerant = 1, facultative anaerobe = 2, obligate aerobe = 3). P values by *t*-test on the regression coefficients, two-tailed. Error bars are 95% C.I. (C) Amino acids in the

IVYWREL group, corresponding to a known signature of thermophile proteomes (Zeldovich *et al.*, 2007), compared to other 13 amino acids by their contribution to the PaPS score (the standardized coefficients from Figure 2). **(D)** IVYWREL content in proteomes of pathogens and non-pathogens. Red lines are medians. **(E,F)** Heat shock resistance in pathogens and non-pathogens. **(E)** survival at 39°C, 41°C, 50°C and 60°C, and **(F)** protein carbonylation at 50°C and 60°C in SHEON, ECOD1, PSEP1, CLOAB, BACC1, LACC3, KLEP3, PSEAE, SALT1, CLOPS, STAAF and STRAG. Error bars represent the standard deviation of three triplicate measurements.

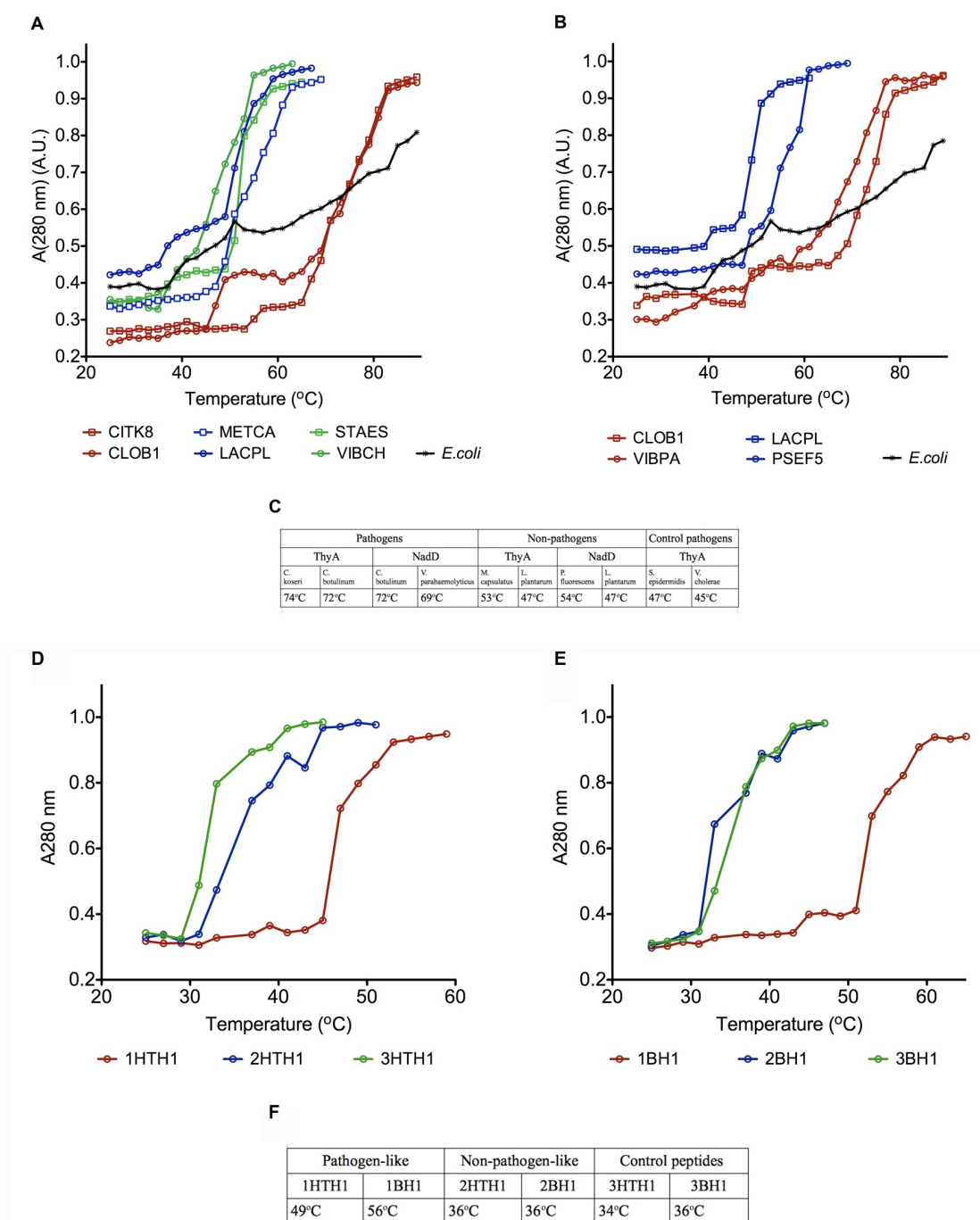


Figure S4, related to Figures 3 and 4. Conformational stability of individual proteins from pathogens and non-pathogens. Thermal denaturation of (A) ThyA from CITK8, CLOB1, METCA, LACPL, STAES and VIBCH, and (B) NadD from CLOB1, VIBPA, LACPL and PSEF5; (C) summary of melting temperatures for each protein. Thermal denaturation of (D) helix-turn-helix (HTH); (E) β -hairpin (BH) designed peptides; (F) summary of melting temperatures for each peptide. Thermal denaturation was monitored as the change of absorbance at 280 nm with increasing temperature. Red symbols are used for pathogen-like proteins and peptides, blue for non-pathogen-like and green for control proteins and peptides. In the name of the peptide, the number in front of the letters denotes (1) a peptide with pathogen-like PaPS, (2) a peptide with non-pathogen-like PaPS, and (3) a negative control peptide (peptides with different overall composition while retaining non-pathogen-like PaPS values). Thermal denaturation of *E.coli* protein extract is added for comparison. The data displays the representative of three independent measurements.

Supplemental Tables

Table S1, relevant throughout the manuscript. Summary of experiments performed on different protein samples with full names and HAMAP codes of species they originate from.

	Full name of the species	HAMAP code of the species
Total protein extract	<i>Pseudomonas aeruginosa</i> <i>Staphylococcus aureus</i> <i>E.coli</i> MG1655 <i>Klebsiella pneumoniae</i> <i>Pseudomonas putida</i> <i>Lactobacillus casei</i> <i>Shewanella oneidensis</i> <i>Bacillus cereus</i> <i>Streptococcus agalactiae</i> <i>Salmonella tify</i> <i>Clostridium perfringens</i> <i>Clostridium acetobutylicum</i>	PSEAE STAAF ECOD1 KLEP3 PSEP1 LACC3 SHEON BACC1 STRAG SALT1 CLOPS CLOAB
ThyA	<i>Lactobacillus plantarum</i> <i>Clostridium botulinum</i> <i>Staphylococcus epidermidis</i> <i>Citrobacter koseri</i> <i>Vibrio cholerae</i> <i>Methylococcus capsulatus</i>	LACPL CLOB1 STAES CITK8 VIBCH METCA
NadD	<i>Lactobacillus plantarum</i> <i>Clostridium botulinum</i> <i>Pseudomonas protegens</i> Pf-5 <i>Vibrio parahaemolyticus</i>	LACPL CLOB1 PSEF5 SHEON

Table S2, related to Figure 2. (presented as a separate file) **A list of analysed bacteria and their phenotypic traits.** “Firmi” stands for “Firmicutes” and “Gamma” for “Gammaproteobacteria”. The “pathogenicity” column shows whether a bacterium is known to be pathogenic to vertebrates, including opportunistic pathogens. Very weakly pathogenic species are marked with asterisks (*). Bacteria pathogenic only to insects are marked with daggers (†). *Shewanella oneidensis* is marked with a question mark (?) to denote that some sources claim it as a possible cause of (rare) infections, while such infections are more commonly attributed to *S. putrefaciens* or *S. algae* in the literature.

Table S3, related to Figure 3C. Melting temperatures of ThyA and NadD in the presence of 1M TMAO.

Non-pathogens				Control pathogens	
ThyA		NadD		ThyA	
M. capsulatus	L. plantarum	P. fluorescens	L. plantarum	S. epidermidis	V. cholerae
68°C	63°C	66°C	62°C	69°C	65°C

Table S4, related to Figure 4. Summary of sequence properties of each synthetic peptide.

Peptide name	PaPS	PUPS	Peptide sequence	Peptide property
1HTH1	-50.84	-57.94	ERELSELGVTRTVVREAIIMLE	Pathogen-like PC1'
2HTH1	51.95	1.73	VSSTAEGVGVSTVSRVIN	Non-pathogen-like PC1'
3HTH1	21.32	-15.26	IKDVAKAAGVGPLREALS	Control peptide
1HTH2	-42.74	-67.47	HRQLAIQYGVTTVVREAIIMLE	Pathogen-like PC1'
2HTH2	38.54	20.13	GLNQIIKESGAPVSTVSRVIN	Non-pathogen-like PC1'
3HTH2	-6.69	-33.21	KLTDVAKLAGVSLREALS	Control peptide
1BH1	-4.15	31.09	DIITLRADKESPKSLVKWESNGVSDY EMKL	Pathogen-like PC1'
2BH1	96.99	105.71	VNWVRQRPQGQGGWIWDIGPGNGTDY N	Non-pathogen-like PC1'
3BH1	53.52	-38.22	GAAVAILGGPGTVWGLAFLDTGPNNAV SAVFLK	Control peptide
1BH2	23.27	-19.74	GTLKFFLTTPVPQGVVRFLQLTVSTGE FKGTVL	Pathogen-like PC1'
2BH2	87.03	85.53	MNWVKQPPGKGGWIWDIGWGDGSR KGVGYVYS	Non-pathogen-like PC1'
3BH2	32.66	-45.89	LAAIATGPGTVSLCVTAPEEFKGTVL	Control peptide