BMC Evolutionary Biology

RESEARCH ARTICLE   **Open Access**

High-throughput quantification of protein structural change reveals potential mechanisms of temperature adaptation in *Mytilus* mussels

Ying-Chen Chao, Melanie Merritt, Devin Schaefferkoetter and Tyler G. Evans*

Abstract

Background: Temperature exerts a strong influence on protein evolution: species living in thermally distinct environments often exhibit adaptive differences in protein structure and function. However, previous research on protein temperature adaptation has focused on small numbers of proteins and on proteins adapted to extreme temperatures. Consequently, less is known about the types and quantity of evolutionary change that occurs to proteins when organisms adapt to small shifts in environmental temperature. In this study, these uncertainties were addressed by developing software that enabled comparison of structural changes associated with temperature adaptation (hydrogen bonding, salt bridge formation, and amino acid use) among large numbers of proteins from warm- and cold-adapted species of marine mussels, *Mytilus galloprovincialis* and *Mytilus trossulus*, respectively.

Results: Small differences in habitat temperature that characterize the evolutionary history of *Mytilus* mussels were sufficient to cause protein structural changes consistent with temperature adaptation. Hydrogen bonds and salt bridges that increase stability and protect against heat-induced denaturation were more abundant in proteins from warm-adapted *M. galloprovincialis* compared with proteins from cold-adapted *M. trossulus*. These structural changes were related to deviations in the use of polar and charged amino acids that facilitate formation of hydrogen bonds and salt bridges within proteins, respectively. Enzymes, in particular those within antioxidant and cell death pathways, were over-represented among proteins with the most hydrogen bonds and salt bridges in warm-adapted *M. galloprovincialis*. Unlike extremophile proteins, temperature adaptation in *Mytilus* proteins did not involve substantial changes in the number of hydrophobic or large volume amino acids, nor in the content of glycine or proline.

Conclusions: Small shifts in organism temperature tolerance, such as that needed to cope with climate warming, may result from structural and functional changes to a small percentage of the proteome. Proteins in which function is dependent on large conformational change, notably enzymes, may be particularly sensitive to temperature perturbation and represent foci for natural selection. Protein temperature adaptation can occur through different types and frequencies of structural change, and adaptive mechanisms used to cope with small shifts in habitat temperature appear different from mechanisms used to retain protein function at temperature extremes.

Keywords: Adaptation, Amino acid, Climate change, Hydrogen bond, Marine, *Mytilus*, Mussel, Protein, Salt bridge, Temperature

* Correspondence: tyler.evans@csueastbay.edu
Department of Biological Sciences, California State University East Bay,
Hayward, CA 94542, USA

Background

Temperature variation, including that resulting from anthropogenic climate change, is a pervasive feature of the biosphere that affects organism fitness through an influence on the three-dimensional structure of proteins [1]. Given that proteins are necessary for the growth, survival, and reproduction of most life on Earth, natural selection should favor amino acid sequences that optimize protein function in particular thermal environments [2–5]. Consistent with this hypothesis, clear patterns of adaptive variation have been discovered in the structural and functional properties of proteins from organisms adapted to different temperatures [5–9]. However, research to date is limited by two issues that have prevented insight into broader-scale patterns of temperature adaptation at the protein level. Firstly, previous work has focused on a small number of proteins, with most studies comparing thermal parameters in a single orthologue between species adapted to different temperatures [10–12]. Because the genomes of multicellular organisms encode thousands of proteins, little inference can be made from existing studies with regard to the proportion of proteins that must be evolutionarily modified in order to increase or decrease organism-level thermal tolerance. Secondly, in instances where the thermal properties of a larger number of proteins have been compared, study systems are typically extremophiles that exhibit very large differences in temperature tolerance and whose proteins are adapted to temperatures outside those encountered by the majority of life [7, 13–19]. Trends resolved in extremophile proteins may not be conserved in multicellular organisms and may not represent mechanisms underlying small adjustments in organism thermal tolerance, for example the few degrees Celsius increase that may be required for persistence in near-future climates [20]. Thus, there exists an opportunity to advance understanding of protein evolution by comparing thermal properties of proteins more comprehensively across the proteome, and doing so in closely related species that differ in thermal tolerance by a small margin.

A fundamental property of proteins is marginal stability: proteins are under selection to maintain a balance between stabilization, which prevents unfolding and aggregation; and flexibility, to allow the shape change necessary for catalysis, substrate binding, or protein-protein interactions [1, 21]. Temperature has a strong influence on the balance between stability and flexibility in proteins. The requirement that proteins have sufficient structural flexibility to undergo conformational change at a given temperature renders proteins susceptible to unfolding when temperature is increased and molecular interactions responsible for stabilizing proteins weaken. Conversely, decreases from optimal temperature cause mobile regions of proteins to become rigid, making shape changes necessary for function more

difficult [5, 22–24]. This trade-off between stability and flexibility ensures no single protein can effectively function over the entire range of temperatures found across the biosphere [5, 25]. Alternatively, evolutionary processes fine-tune protein structures such that function is optimized at temperatures most frequently encountered by cells: orthologous proteins exhibit differences in thermal stability such that an ortholog from a cold-adapted species is more thermally labile than an ortholog from a warm-adapted species [3, 26–28].

Comparison of proteins adapted to different temperatures demonstrates that there are a multitude of ways in which thermal performance can be adaptively modified [29–31]. Altering hydrophobicity, charge, non-covalent interactions, volume, and cooperativity have all been implicated in protein temperature adaptation [14, 25, 32–36]. Also apparent from these comparisons is that the location of adaptive change within the folded protein is important. Natural selection can preferentially act upon particular regions during protein temperature adaption, such as the hydrophobic core or solvent-exposed surface, or target specific domains, such as coils, loops, helixes, or substrate binding sites, to adaptively shift thermal stability [5, 36, 37]. Often these mechanisms of thermal adaptation are reflected in specific patterns of amino acid use within proteins [6, 7, 9, 37]. Amino acid substitutions that adaptively vary the number of non-covalent hydrogen bonds and salt-bridges in proteins, including addition or removal of polar and charged residues, are a common means of optimizing protein function at given environmental temperatures [6]. The number of hydrogen bonds and salt bridges tends to increase in heat-tolerant (thermophilic) proteins as a means of protecting against denaturation at high temperatures [32, 33, 38–40]. Conversely, these stabilizing features tend to decrease in cold-tolerant (psychrophilic) proteins so that proteins remain sufficiently flexible to undergo conformational change at low temperatures [41–43]. Thermal stability is also modulated by the presence of hydrophobic and large volume amino acids. Residues with these chemical properties enhance thermal stability by restricting solvent access to the protein core [35], and accordingly, are frequently found in greater abundance in proteins that retain function in the heat [6, 34]. The ability of cysteine residues to form disulfide bonds, the rigidity of proline compared with other amino acids, and the lack of a side chain on glycine, also results in content of these amino acids often being correlated with protein structural stability and organism temperature tolerance [4].

In this study, software was developed to quantify hydrogen bonds, salt bridges, and amino acids across the proteomes of two closely related species of marine mussel that differ in temperature tolerance by a few

Chao *et al. BMC Evolutionary Biology*   (2020) 20:28

Page 3 of 18

degrees Celsius: *Mytilus galloprovincialis* and *Mytilus trossulus* [44]. This methodological innovation enabled comparison of structural features among large numbers of *Mytilus* proteins, and in turn, broader inference into how small shifts in environmental temperature influence the structural properties of proteins. *Mytilus* mussels have proven excellent model systems to examine protein temperature adaptation [44–46]. Although closely related, *M. galloprovincialis* and *M. trossulus* evolved in distinct thermal environments. *M. trossulus* and *M. galloprovincialis* are members of the 'blue mussel complex', which also includes *M. edulis* [47]. *M. trossulus*, a native of the cold Northeast Pacific Ocean, appears to be the ancestral species of the complex, with divergence into *M. edulis* occurring as a result of migration through a transient opening in the Bering Strait and into the Atlantic Ocean approximately 3.5 million years ago [48]. Subsequently, *M. edulis* expanded its range throughout the North Atlantic and into the warmer waters of the Mediterranean Sea, where populations became isolated and gave rise to *M. galloprovincialis* about 2 million years ago [49]. Differences in prevailing temperatures experienced during evolutionary history are reflected in the thermal tolerances of *M. galloprovincialis* and *M. trossulus*: the upper critical temperature of *M. galloprovincialis* is approximately 4 °C higher than that of *M. trossulus* as a result of evolution in a comparatively warmer environment (LT$_{50}$ = 38 °C and LT$_{50}$ = 34 °C, respectively) [12]. Physiological performance also varies between these species in a manner that is consistent with adaptation to different temperatures. All physiological (e.g. heart rate [50]), biochemical (e.g. enzyme activities [44–46]), and molecular (e.g. gene expression [51, 52];) comparisons made to date show *M. galloprovincialis* to be more tolerant of heat and *M. trossulus* to be more tolerant of cold.

Orthologous enzymes from *M. galloprovincialis* and *M. trossulus* exhibit differences in structural stability and function that are congruent with their thermal tolerances and with known mechanisms of temperature adaptation. Cytosolic malate dehydrogenase (cMDH) and isocitrate dehydrogenase (IDH) from warm-adapted *M. galloprovincialis* sustain activity and optimal substrate binding at higher temperatures than orthologs from cold-adapted *M. trossulus* [45, 46]. Enhanced function of *M. galloprovincialis* cMDH and IDH results from amino acid substitutions that increase the number of hydrogen bonds in order to protect regions undergoing conformational change from unfolding at high temperatures [3, 45, 46]. However, the present state of research provides an incomplete perspective about how natural selection modifies *Mytilus* proteins to optimize function in their respective temperature regimes. Only a small number of *Mytilus* proteins have been analyzed: the

influence of temperature on protein function has been investigated in only six *Mytilus* orthologs to date [45, 46]. Thermal stability has been predicted in silico for large numbers of *Mytilus* proteins, and importantly, this work has confirmed the presence of more stabilizing amino acids in the proteins of *M. galloprovincialis* compared with proteins from *M. trossulus* [53]. However, the methodology used in this previous study did not allow for inference into the potential contributions of hydrogen bonds or salt bridges to variation in protein thermal stability, nor could it identify the specific types of amino acid substitution responsible for differences in protein thermal stability between *Mytilus* congeners.

Here, individual amino acids, hydrogen bonds, and salt bridges were counted in large numbers of *M. galloprovincialis* and *M. trossulus* proteins in order to address several unresolved questions surrounding protein evolution in *Mytilus* and protein temperature adaptation in general [1, 5, 54]. First, are small differences in environmental temperature sufficient to cause changes in amino acid use and non-covalent interactions that are consistent with mechanisms of protein temperature adaptation? Second, what proportion of the proteome needs to be adaptively modified to alter organism temperature tolerance by a small margin? Third, are proteins with particular functions more likely than others to be acted on by natural selection during temperature adaptation? Finally, are certain mechanisms of adaptive change favored among *Mytilus* proteins and do these changes parallel those occurring in proteins adapted to temperature extremes? Answers to these questions have broad implications, from improving estimates of the capacity to cope with climate change, to better understanding the process of protein evolution over the range of temperatures found throughout the biosphere.

Methods

Amino acid sequences, protein modeling, and quantifying hydrogen bonds and salt bridges

Hydrogen bonds or salt bridges are rarely directly measured within proteins [55]. Instead, these structural features are predicted based on the spatial arrangement of atoms within three-dimensional protein models created from amino acid sequences. *M. galloprovincialis* and *M. trossulus* protein sequences were obtained by searching for the keyword "*Mytilus*" in the UniProt database [56] and then selecting all sequences for each species using the "popular organisms" tab on the results page. Resulting sequences were downloaded in .fasta format and used to create models of each protein using SWISS MODEL [57, 58]. SWISS MODEL software uses sequence information to search a database of proteins for which the three-dimensional structure has been solved using X-ray crystallography, nuclear magnetic resonance,

or electron microscopy techniques [59]. When an appropriate template has been identified, the software develops a model of the protein of interest, storing information about the location of each amino acid in a protein database (.pdb) file. SWISS MODEL will model proteins in the macromolecular assembly that has either been shown to be or is believed to be the functional form of the molecule. SWISS MODEL templates are annotated with information about the whether the protein functions as a monomer or an oligomer (referred to as the biological assembly or biounit), and this quaternary structure annotation is used to model the target sequence in its functional form [58]. SWISS MODEL will include a ligand in protein models if the highest scoring template includes a biologically relevant ligand and amino acid residues involved in ligand binding are conserved in the target–template sequence alignment [58].

Mytilus .pdb files generated by SWISS MODEL were uploaded to a web server of the HBOND program that uses the spatial arrangement of amino acids recorded in the .pdb file to predict hydrogen bonding [60]. The HBOND algorithm predicts hydrogen bond formation at a maximum donor–acceptor distance of 3.5 Å and maximum hydrogen–acceptor distance of 2.5 Å. HBOND does not consider bond angle in its prediction criteria. HBOND will calculate hydrogen bonds between a protein and ligand if a ligand is included in the model, but does not consider hydrogen bonds that could form between the protein and solvent [60]. The same .pdb files were also uploaded to a different web server called Evaluating Salt BRIdges in proteins (ESBRI) in order to predict salt bridges in each protein [61]. The ESBRI algorithm predicts formation of a salt bridge if at least one aspartate or glutamate side-chain carboxyl oxygen atom and one side-chain nitrogen atom of arginine, lysine, or histidine are within a distance of 4.0 Å. Predicted hydrogen bonds and salt bridges were then counted from the HBOND and ESBRI output files, respectively, to get the total number of each feature in each *Mytilus* protein of interest.

Software development

Manual completion of the protocol described above constrains the number of proteins that can be assayed. To allow high-throughput quantification of hydrogen bonds and salt bridges in proteins of interest, custom software was developed that automated the data acquisition pipeline. The software required as input a .fasta file containing all proteins of interest. Individual sequences from the .fasta file were then sequentially passed to SWISS MODEL in order to generate protein models. Resulting .pdb files were then automatically uploaded to HBOND and ESBRI servers in order to predict hydrogen bonds and salt bridges. Pertinent information was extracted

from output files at each step of the process, including the unique UniProt ID and length of each protein (extracted from the .fasta file), QMEAN (Qualitative Model Energy ANalysis, a quality control metric extracted from the .pdb file), and the total number of hydrogen bonds and salt bridges (extracted from the HBOND and ESBRI output files). These parameters were saved to a database until data from all proteins in the .fasta input file had been processed by the software.

Quality control and statistics

QMEAN is a score generated by SWISS MODEL to estimate the quality of protein models. QMEAN is a combined scoring function that incorporates five different structural descriptors to determine the major geometrical aspects of protein structure, with a higher score indicative of greater agreement between the protein model and the native conformation of the template protein [62]. QMEAN score was used to remove low quality models of *Mytilus* proteins that could cause inaccurate prediction of hydrogen bonds and salt bridges. Following SWISS MODEL recommendations, protein models with QMEAN scores less than – 5 were removed from analyses [63]; personal communication, Andrew Waterhouse, Swiss Institute of Bioinformatics]. Duplicate proteins within each species were removed by retaining the single protein model with the highest QMEAN score. This filtering step was necessary because the UniProt database may contain multiple entries for the same protein (i.e. proteins annotated exactly the same, but assigned a different UniProt ID) and because SWISS MODEL often generates multiple models for the same protein using different templates. Proteins annotated as "uncharacterized protein" may not represent unique proteins and were removed from analyses. To control for differences in length when comparing *Mytilus* proteins (long proteins composed of many amino acids would be expected to have more hydrogen bonds and salt bridges than short proteins), the number of hydrogen bonds and the number of salt bridges were each divided by the total number of amino acids in each protein. The resulting values, expressed as the number of hydrogen bonds or salt bridges per amino acid for each unique protein, were then used in statistical tests comparing sets of proteins between the two mussel species. Proteins annotated exactly the same between each species were considered orthologs.

Anderson-Darling, Ryan-Joiner, and Kolmogorov-Smirnov tests of normality all indicated that the number of hydrogen bonds per amino acid and the number of salt bridges per amino acid in *Mytilus* proteins were not normally distributed. Standard methods of data transformation (e.g. log, square, and Box-Cox transformations) did not correct this trend. Given these results,

Chao *et al. BMC Evolutionary Biology*   (2020) 20:28

Page 5 of 18

non-parametric Mann-Whitney U tests were used to determine whether the median number of hydrogen bonds or salt bridges per amino acid significantly differed between *M. galloprovincialis* and *M. trossulus*. Statistical tests were performed using MiniTab version 18.0 with *p*-values less than 0.05 considered significant. Box plots and frequency histograms used to illustrate hydrogen bond and salt bridge data were created using Microsoft Excel 2016.

Over-representation of protein ontologies
Over-representation analysis was used to determine if *Mytilus* proteins with particular functions were more likely than others to have modified the number of hydrogen bonds or salt bridges. Over-representation analysis is a statistical approach that uses functional information to identify categories of proteins (called ontologies [64]) found in a greater proportion than expected within a user-defined list. Significance is determined by the probability that the number of proteins from a given ontology in the user-defined list occurred by chance relative to the number of proteins from this same ontology in a larger background list. Over-representation analysis was performed using the R package TopGO using the "classic" algorithm in which each ontology category from the biological process, molecular function, and cellular component databases are tested independently [65]. Background lists were created for each species by including all unique proteins with QMEAN scores greater than − 5. Ontologies with unadjusted Fisher Exact test *p*-values less than 0.05 were considered significantly over-represented (correction for multiple testing can produce misleading results and is not recommended in TopGO [66]).

Amino acid use
Custom software used to quantify hydrogen bonds and salt bridges was also programmed to count the number of each amino acid in each protein from the input .fasta file. Chi-square tests of independence were then used to determine if the proportions of each amino acid differed among sets of *M. galloprovincialis* and *M. trossulus* proteins, with *p*-values less than 0.05 considered significant [67]. This statistical test generates observed and expected values for each amino acid for each species; thus, differences in these two values provide insight into amino acid use and potential mechanisms of temperature adaptation among *Mytilus* proteins. Observed values represent total counts of each amino acid within the proteins from each species. Expected values are calculated by determining the frequency of each amino acid among proteins from both species in a given protein set and then multiplying this frequency by the actual observed count of each amino acid in the proteins

from each species individually in the set. Amino acid use was first analyzed in subsets of proteins with the greatest number of hydrogen bonds and salt bridges per amino acid so that specific amino acid substitution patterns could be related to differences in the number of these structural features between *M. galloprovincialis* and *M. trossulus*. Amino acid content of all unique proteins with sequence information for each species (including protein sequences from which SWISS-MODEL could not create a high quality model and therefore hydrogen bonds and salt bridges could not be quantified) was also analyzed using the same statistical approach. Analyzing all proteins provided a broader overview of amino acid use among *Mytilus* proteins.

Amino acid data were also used to test the hypothesis that the proportion of hydrophobic and large volume amino acids would be greater among the proteins of warm-adapted *M. galloprovincialis* compared with the proteins of cold-adapted *M. trossulus*. This hypothesis was statistically evaluated by calculating Spearman rank order correlation coefficients (rho) between the deviation (i.e. observed-expected) of each amino acid in *M. galloprovincialis* relative to *M. trossulus* and hydrophobicity rank [68] or volume rank [69] for each residue. All unique proteins with amino acid sequence information from each species were used in this analysis. These same tests were performed on proteins with the greatest number of hydrogen bonds or salt bridges per amino acid to determine if changes in hydrophobicity or residue volume were occurring in the same subset of proteins exhibiting differences in hydrogen bonds and salt bridges between *Mytilus* species.

Results
Protein sequences, modeling, and quality control
The UniProt database contained 4609 protein sequences for *M. galloprovincialis* and 725 protein sequences for *M. trossulus* (Table 1). Three-dimensional modeling of these 5334 protein sequences resulted in 1325 models generated for *M. galloprovincialis* and 669 models for *M. trossulus*. An X in the input sequence (indicating a missing or ambiguous amino acid) or absence of a suitable template protein in the SWISS MODEL database caused the number of protein models generated to be smaller than the number of input sequences. Removal of models for duplicate proteins within each species (for example Cytochrome C oxidase 1 accounted for 137 and 228 models in *M. galloprovincialis* and *M. trossulus*, respectively), left 574 unique protein models in *M. galloprovincialis* and 85 unique protein models in *M. trossulus*. Of the remaining models, 482 and 58 had QMEAN scores greater than − 5 in *M. galloprovincialis* and *M. trossulus*, respectively, and were used in subsequent analyses of hydrogen bonding and salt bridge

Table 1 Summary of protein sequences, protein modeling, and filtering

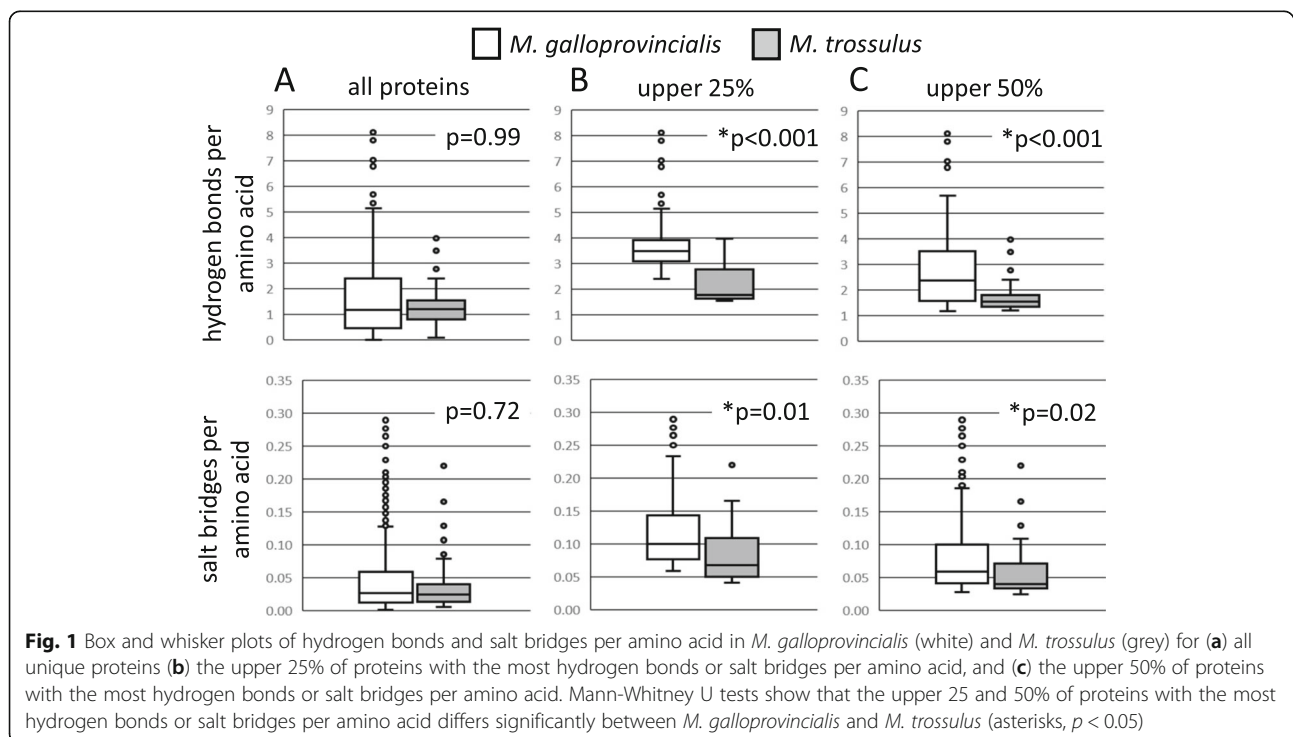|	*M. galloprovincialis*	*M. trossulus*
Number of proteins in UniProt database	4609	725
Number of protein models generated in SWISS MODEL	1325	669
Number of models for unique proteins	574	85
Number of models for unique proteins with QMEAN > −5	482	58
Number of orthologous proteins	19	19

formation (Additional file 1: Figures S1 and S2). Within this set of high-quality models, 19 proteins were annotated exactly the same between *M. galloprovincialis* and *M. trossulus* and were presumed to be orthologs (Additional file 1: Figure S3).
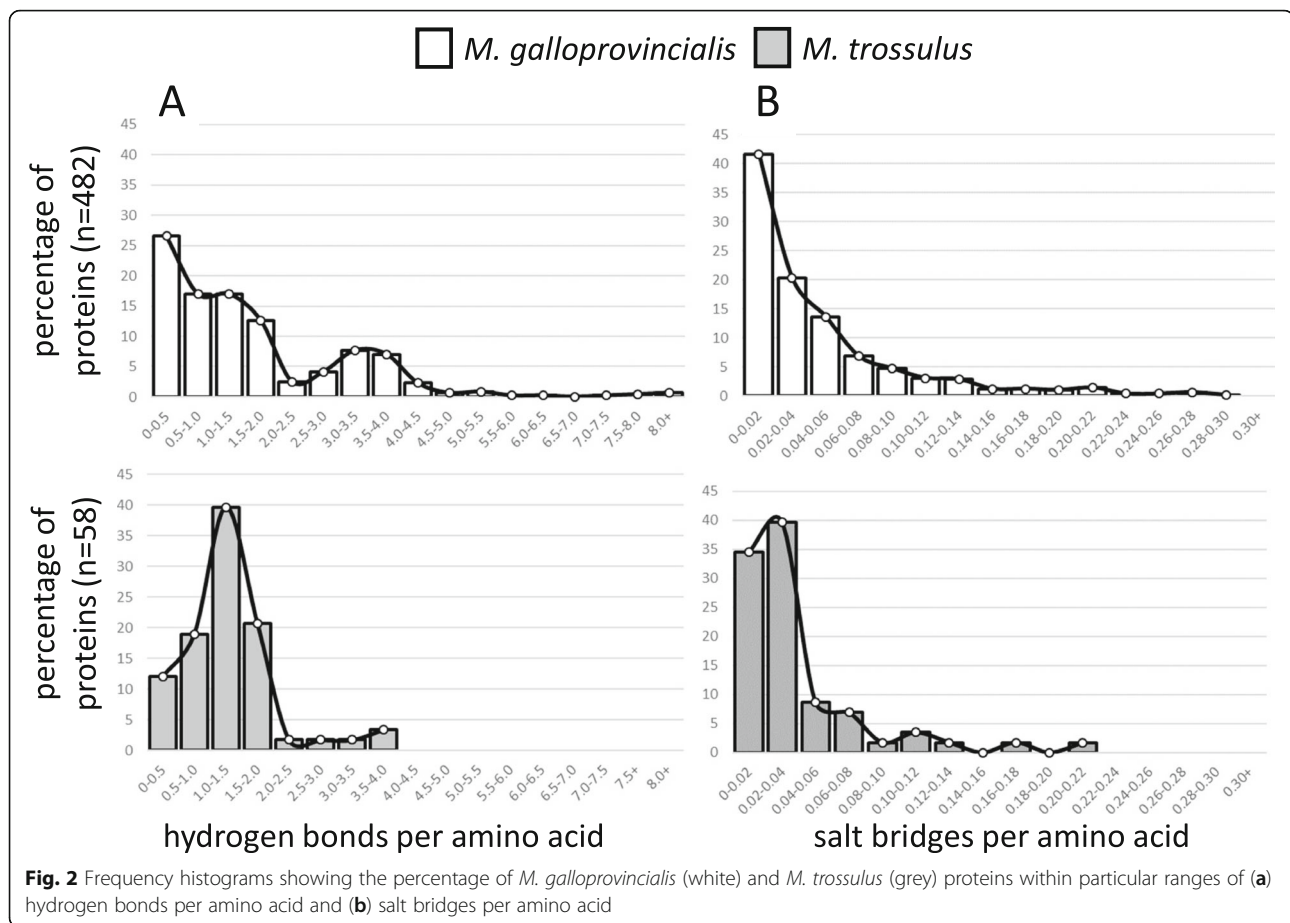
Hydrogen bonds and salt bridges

The number of hydrogen bonds and salt bridges per amino acid were compared to test the hypothesis that these stabilizing features are more common within the proteins of warm-adapted *M. galloprovincialis* than within the proteins of cold-adapted *M. trossulus*. The median number of hydrogen bonds or salt bridges per amino acid among all unique proteins did not differ significantly between *M. galloprovincialis* and *M. trossulus* (Fig. 1a). There was also no significant difference between the two species in the median number of hydrogen bonds or salt bridges per amino acid in the set of 19 orthologous proteins. Given these results, an alternative hypothesis is that enhanced heat tolerance in *M. galloprovincialis* is the result of added stability in a smaller

subset of proteins [46]. This hypothesis was tested by generating frequency histograms that compare the percentage of proteins within different ranges of hydrogen bonds and salt bridges per amino acid between the two species (Fig. 2). These histograms show that *M. galloprovincialis* has a greater percentage of proteins within the upper range of hydrogen bonds and salt bridges per amino acid compared with *M. trossulus*. For example, 5.4% (26 of 482) of *M. galloprovincialis* proteins have more than 4.0 hydrogen bonds per amino acid, whereas none of the 58 *M. trossulus* proteins analyzed meet this criteria (Fig. 2a). Similarly, 6.6% (32 of 482) of *M. galloprovincialis* proteins have more than 0.14 salt bridges per amino acid compared with only 3.4% (2 of 58) of *M. trossulus* proteins (Fig. 2b).

Frequency histograms indicate that the number of hydrogen bonds and salt bridges differ in a subset of *Mytilus* proteins. This possibility was further evaluated by statistically comparing the median number of hydrogen bonds and salt bridges per amino acid in the upper quartiles and upper halves of proteins in each species.

Fig. 1 Box and whisker plots of hydrogen bonds and salt bridges per amino acid in *M. galloprovincialis* (white) and *M. trossulus* (grey) for (**a**) all unique proteins (**b**) the upper 25% of proteins with the most hydrogen bonds or salt bridges per amino acid, and (**c**) the upper 50% of proteins with the most hydrogen bonds or salt bridges per amino acid. Mann-Whitney U tests show that the upper 25 and 50% of proteins with the most hydrogen bonds or salt bridges per amino acid differs significantly between *M. galloprovincialis* and *M. trossulus* (asterisks, *p* < 0.05)

Fig. 2 Frequency histograms showing the percentage of *M. galloprovincialis* (white) and *M. trossulus* (grey) proteins within particular ranges of (**a**) hydrogen bonds per amino acid and (**b**) salt bridges per amino acid

The upper 25% of *M. galloprovincialis* proteins ($n = 121$) contained significantly more hydrogen bonds ($p < 0.001$) and salt bridges ($p = 0.011$) per amino acid than did the upper 25% of *M. trossulus* proteins ($n = 15$) (Fig. 1b). This pattern also extended to the upper 50% of proteins, with *M. galloprovincialis* ($n = 241$) having significantly more hydrogen bonds ($p < 0.001$) and salt bridges ($p = 0.024$) per amino acid compared with *M. trossulus* ($n = 29$) (Fig. 1c).

Over-representation of protein ontologies

Over-representation analysis was performed on the upper 25% and upper 50% of *M. galloprovincialis* proteins to determine if potentially adaptive increases in hydrogen bonding and salt bridge formation occurred disproportionately in proteins within certain pathways, in proteins with particular functions, or in proteins at specific cellular locations (Additional file 1: Figure S4). The upper 25% of *M. galloprovincialis* proteins with the greatest number of hydrogen bonds per amino acid contained a significantly greater proportion of proteins involved in carboxylic acid metabolism (e.g. "tricarboxylic acid cycle"), programmed cell death (e.g. "apoptotic process"), and antioxidant production (e.g. "glutathione

transferase activity") (Table 2). Proteins annotating to the carboxylic acid metabolism ontology included cMDH and IDH. Proteins annotating to cell death ontologies included regulatory molecules within the apoptosis signaling cascade (e.g. multiple isoforms of p63) and proteases involved in the destruction of proteins (e.g. caspase 3/7−1 and 3/7−2) [70–73]. Six isoforms of glutathione-S-transferase enzymes involved in the production of antioxidants were present in the upper 25% of *M. galloprovincialis* proteins with the most hydrogen bonds per amino acid [74]. Ontologies relating to these same three processes were also over-represented in the upper 50% of *M. galloprovincialis* proteins with the most hydrogen bonds per amino acid, including "oxidation-reduction process"(carboxylic acid metabolism), "cysteine-type peptidase activity" (cell death), and "transferase activity" (antioxidant production). Cellular component ontology terms were not significantly over-represented in either the upper 25% or upper 50% of *M. galloprovincialis* proteins with the greatest number of hydrogen bonds per amino acid (Additional file 1: Figure S4).

In general, the same ontologies that were significantly over-represented among sets of proteins with the most hydrogen bonds per amino acid were also over-represented

Table 2 Select ontology over-representation among the 25% of *M. galloprovincialis* proteins with most hydrogen bonds per amino acid

Gene ontology ID	Ontology	Number in background	Number in upper 25%	Number expected in upper 25%	*P*-value
A. Biological process					
GO:0006082	organic acid metabolic process	3	3	0.38	0.001
GO:0006091	carboxylic acid metabolic process	3	3	0.38	0.001
GO:0006099	tricarboxylic acid cycle	2	2	0.25	0.014
GO:0006915	apoptotic process	15	6	1.88	0.001
GO:0008219	cell death	15	6	1.88	0.001
B. Molecular function					
GO:0016740	transferase activity	10	7	1.44	< 0.001
GO:0004364	glutathione transferase activity	3	3	0.43	0.003

among sets of proteins with the most salt bridges per amino acid, including those relating to carboxylic acid metabolism (e.g. "organic acid metabolic process"), cell death (e.g. "apoptotic process"), and antioxidant production (e.g. "glutathione transferase activity") (Table 3; Additional file 1: Figure S4). Overlap between proteins with the most hydrogen bonds and those with the most salt bridges was not unexpected since salt bridges contain a hydrogen bond. Consistent with this assumption, 71% of proteins (*n* = 200) were found in both the list of the 50% of proteins with the greatest number of hydrogen bonds per amino acid and the list of the 50% of proteins with the greatest number of salt bridges per amino acid in *M. galloprovincialis*. Spearman's rank tests show that the number of hydrogen bonds per amino acid was significantly positively correlated with the number of salt bridges per amino acid for all proteins, as well as the upper 50 and 25% of proteins with the most hydrogens bonds or salt bridges per amino acid for both species (Additional file 1: Figure S5).

Amino acid use

Amino acids differ in their ability to form hydrogen bonds and salt bridges, thus certain amino acids were expected to be more frequent in *M. galloprovincialis* proteins that exhibited higher rates of hydrogen bonding and salt bridge formation compared with *M. trossulus* (Additional file 1:

Figures S6 and S7). A chi square test for independence demonstrates that amino acid use was significantly different between *M. galloprovincialis* and *M. trossulus* in both the upper 25% and upper 50% of proteins with the greatest number of hydrogen bonds per amino acid. Amino acids with hydroxyl (serine and threonine), amide (asparagine and glutamine), aromatic (histidine, tyrosine, tryptophan), or charged groups (aspartate, glutamate, arginine, lysine), are most likely to form conventional (OH--O or OH--N) hydrogen bonds [75]. The upper 50% of *M. galloprovincialis* proteins with the most hydrogen bonds per amino acid contained 105 more of these 11 amino acids than expected compared with the upper 50% of *M. trossulus* proteins (1.4% higher use rate; Table 4). Similarly, the upper 25% of *M. galloprovincialis* proteins contained 51 more of these 11 amino acids than expected compared with the upper 25% of *M. trossulus* proteins (1.2% higher use rate; Additional file 1: Figure S8). These trends were caused by increases in the number of aspartate, asparagine, glutamine, and threonine residues in both the upper 25 and 50% of *M. galloprovincialis* proteins. Amino acid use did not differ significantly between *M. galloprovincialis* and *M. trossulus* among the set of 19 presumed orthologs (Additional file 1: Figure S8).

Salt bridges form as a result of attraction between oppositely charged residues; thus, the proportions of amino

Table 3 Select ontology over-representation among the 25% of *M. galloprovincialis* proteins with most salt bridges per amino acid

Gene ontology ID	Ontology	Number in background	Number in upper 25%	Number expected in upper 25%	*P*-value
A. Biological process					
GO:0006082	organic acid metabolic process	3	3	0.83	0.019
GO:0019752	carboxylic acid metabolic process	3	3	0.83	0.019
GO:0006915	apoptotic process	15	8	4.17	0.018
GO:0008219	cell death	15	8	4.17	0.018
B. Molecular function					
GO:0016740	transferase activity	10	8	2.80	< 0.001
GO:0004364	glutathione transferase activity	3	3	0.84	0.020

Chao et al. BMC Evolutionary Biology (2020) 20:28

Page 9 of 18

Table 4 Amino acid use among the 50% of *Mytilus* proteins with the most hydrogen bonds per amino acid. Amino acids most likely to form conventional hydrogen bonds are shaded grey [75]

							Mytilus galloprovincialis													
Amino Acid	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Iso	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Observed	3888	2944	3415	3704	1329	2636	3512	4760	1802	3976	5523	4177	1797	2979	2676	4899	4237	825	2448	4535
Expected	3934	3006	3343	3605	1318	2604	3503	4755	1803	3952	5489	4268	1850	2973	2753	4899	4166	854	2443	4544
Difference	-46	-62	+72	+99	+11	+32	+9	+5	-1	+24	+35	-91	-53	+6	-77	0	+72	-29	+5	-9
Percent Usage	5.9%	4.5%	5.2%	5.6%	2.0%	4.0%	5.3%	7.2%	2.7%	6.0%	8.4%	6.3%	2.7%	4.5%	4.1%	7.4%	6.4%	1.2%	3.7%	6.9%
							Mytilus trossulus													
Amino Acid	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Iso	Leu	Lys	Met	Phe	Pro	Ser	Thr	Try	Tyr	Val
Observed	537	437	345	351	153	293	428	588	226	469	650	623	284	365	420	611	448	135	300	576
Expected	491	375	417	450	164	325	437	593	225	493	685	532	231	371	343	611	520	107	305	566.7
Difference	46	+62	-72	-99	-11	-32	-9	-5	+1	-24	-35	+91	+53	-6	+77	0	-72	+29	-5	+9
Percent Usage	6.5%	5.3%	4.2%	4.3%	1.9%	3.6%	5.2%	7.1%	2.7%	5.7%	7.9%	7.6%	3.4%	4.4%	5.1%	7.4%	5.4%	1.6%	3.6%	7.0%

acids with charged side chains (arginine, aspartate, lysine, and glutamate) were predicted to differ between the proteins of *M. galloprovincialis* and *M. trossulus* (Additional file 1: Figures S6 and S7). The upper 25% and upper 50% of proteins with the greatest number of salt bridges per amino acid differed significantly in amino acid composition between *Mytilus* species (Additional file 1: Figure S8). Among the 50% of proteins with the most salt bridges per amino acid, there were 80 more negatively charged residues than expected in *M. galloprovincialis* compared with *M. trossulus*, a trend caused by an increase in the number of aspartate residues (1.2% higher use rate; Table 5). The number of positively charged amino acids decreased in *M. galloprovincialis* proteins relative to *M. trossulus*. The 50% of *M. galloprovincialis* proteins with the most salt bridges per amino acid had 132 fewer lysine residues (2.0% lower use rate) and 100 fewer arginine residues than expected (1.5% lower use rate) compared with *M. trossulus*. Inclusion of histidine, which is positively charged at physiological pH, has little effect on these

trends because the number of histidine residues had very small deviations from the expected number between species (Table 5).

Analysis of thermophilic proteins from prokaryotes indicates that stability and function in the heat can be improved by increasing the number of disulfide bridges, by reducing glycine content, by increasing hydrophobicity of the protein core, and by increasing protein volume [1, 6, 76]. Deviations in amino acid use were used to determine whether these same mechanisms of protein temperature adaptation contribute to protein evolution in *Mytilus* mussels. Amino acid use provides support for some, but not all, of these changes occurring in *Mytilus* proteins (Additional file 1: Figure S8). The complete set of *M. galloprovincialis* proteins had more cysteines than expected compared with all *M. trossulus* proteins (+ 109; 0.6% higher in use rate), similar to the amino acid composition of thermophilic proteins. However, glycine content increased in warm-adapted *M. galloprovincialis* relative to *M. trossulus* (+ 39; 0.2% higher use rate), a

Table 5 Amino acid use among the 50% of *Mytilus* proteins with the most salt bridges per amino acid. Positively charged residues are shaded black. Negatively charged residues are shaded grey

							Mytilus galloprovincialis													
Amino Acid	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Iso	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Observed	4324	3485	3804	4366	1516	3120	4408	5099	1932	4411	6071	4942	1947	3068	3178	5353	4559	837	2591	4838
Expected	4358	3585	3748	4286	1498	3114	4409	5077	1930	4385	5986	5074	1965	3018	3251	5359	4553	833	2589	4833
Difference	-34	-100	+56	+81	+18	+6	-1	+22	+2	+26	+85	-132	-18	+50	-73	-6	+6	+4	+2	+6
Percent Usage	5.9%	4.7%	5.2%	5.9%	2.1%	4.2%	6.0%	6.9%	2.6%	6.0%	8.2%	6.7%	2.6%	4.2%	4.3%	7.2%	6.2%	1.1%	3.5%	6.6%
							Mytilus trossulus													
Amino Acid	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Iso	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Observed	472	460	321	350	132	307	444	488	192	415	516	642	215	253	399	544	451	80	258	480
Expected	438	360	377	431	150	313	443	510	194	441	601	510	197	303	327	538	457	84	260	486
Difference	+34	+100	-56	-81	-18	-6	+1	-22	-2	-26	-85	+132	+18	-50	+73	+6	-6	-4	-2	-6
Percent Usage	6.4%	6.2%	4.3%	4.7%	1.8%	4.1%	6.0%	6.6%	2.6%	5.6%	7.0%	8.7%	2.9%	3.4%	5.4%	7.3%	6.1%	1.1%	3.5%	6.5%

result that does not conform to that of thermophilic proteins. Also in opposition to patterns observed in thermophiles, *M. galloprovincialis* did not increase the proportion of hydrophobic or large volume amino acids relative to *M. trossulus* when all proteins are considered (Fig. 3). There is also no significant positive correlation between the proportion of hydrophobic or large volume amino acids in the upper 25% or upper 50% of *Mytilus* proteins with either the highest number of hydrogen bonds per amino acid or the highest number of salt bridges per amino acid (Additional file 1: Figure S9).

Discussion

The study of protein evolution presently benefits from a vast amount of sequence data, but a relatively small body of structural data [24]. Software developed and applied here addresses this deficiency by capitalizing on existing amino acid sequence information to generate new protein structural data. Liao et al. (2017) emphasize that effects of a single substitution cannot be fully interpreted by examining only localized bonding in specific amino acid residues, but must include a broader analysis of changes in protein structure [10]. Here, for the first time, the total number of hydrogen bonds, salt bridges, and amino acids in large numbers of *Mytilus* proteins were analyzed to reveal broader patterns of structural change between species. Fields et al. (2015) iterate two unresolved issues

regarding protein temperature adaptation [5]. One, that we are currently unable to place comparative studies of protein evolution into the broader context of the proteome because too few proteins have been analyzed. And two, the need to identify protein classes or biochemical pathways that are more frequently acted upon by natural selection during temperature adaptation. Analyses performed here suggest that comprehensive modification of the proteome is not required to alter thermal tolerance by a small margin because only a small percentage of *Mytilus* proteins showed structural variation between species consistent with temperature adaptation. Over-representation analysis indicated that enzymes, which require large conformation change during catalysis, were disproportionately acted on by natural selection during protein temperature adaptation in *Mytilus*. Finally, it remains important to evaluate whether a given evolutionary solution for sustaining protein function at different temperatures is conserved across phylogenetic boundaries [1, 12, 77]. Structural changes associated with temperature adaptation in *Mytilus* mussels were contrasted with mechanisms established in extremophile proteins to address this issue.

Variation in hydrogen bonding and salt bridge formation among *Mytilus* proteins

Moderately higher upper critical temperature of *M. galloprovincialis* compared with *M. trossulus* was associated
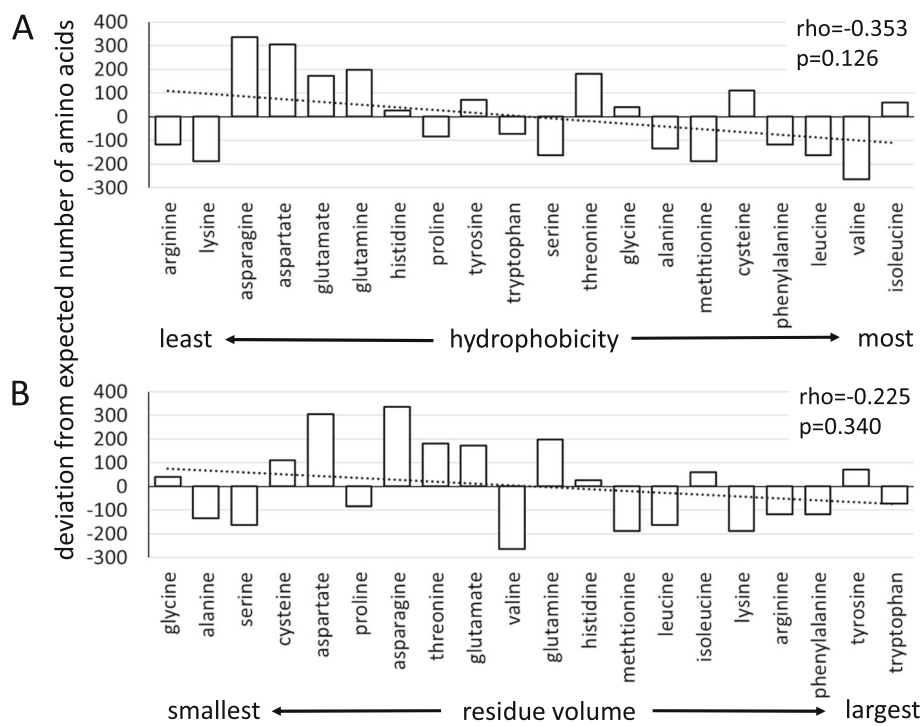

Fig. 3 Spearman rank order correlation between the deviation in amino acid use in *M. galloprovincialis* relative to *M. trossulus* for all proteins and (**a**) rank residue hydrophobicity or (**b**) rank residue volume

with higher numbers of stabilizing hydrogen bonds and salt bridges in a small percentage of proteins. A broad interpretation of this result is that small increases in organism thermal tolerance may be achieved through evolutionary modification of only part of the proteome. Other analyses of *Mytilus* proteins substantiate this interpretation, demonstrating that structural or functional changes that accompany temperature adaptation occur in only a portion of the proteins analyzed. The effect of temperature on the kinetic properties of six *Mytilus* enzyme othologs have been assessed to date, and in only two, cMDH and IDH, did Michaelis-Menten (Km) values differ significantly between *M. galloprovincialis* and *M. trossulus* [45, 46]. Thermal sensitivities of the other four orthologs were remarkably similar between *Mytilus* species over a wide range of temperatures [46]. In silico estimates of stability differences among *Mytilus* proteins further corroborate that evolutionary processes underlying temperature adaptation in these species favor changes occurring in particular proteins rather than changes occurring widely across the proteome [53]. A significantly greater number of stabilizing amino acid substitutions occurred in the proteins of warm-adapted *M. galloprovincialis* compared with cold-adapted *M. trossulus*. However, this signal of thermal adaptation was caused by a small number of *M. galloprovincialis* outlier proteins in which amino acid changes resulted in large increases in protein stability. Importantly, removal of these outlier proteins from statistical models caused differences in average protein thermal stability between the two species to become insignificant [53]. Even very large differences in organism temperature tolerance do not appear to result from comprehensive modification of the proteome. Comparison of extremophile Archaea and bacteria adapted to temperatures that range from − 12 °C to 103 °C indicates that protein thermal stability is modified to different extents across the proteome. Stability was increased in many thermophile proteins as expected; however, other proteins exhibited no difference in thermal stability between thermophiles and psychrophiles, and some proteins exhibited greater stability in cold-tolerant species contrary to a priori predictions [19].

Functional bias in *Mytilus* protein temperature adaptation

Significant differences in hydrogen bonding and salt bridge formation occurring among a subset of *Mytilus* proteins indicates that the strength of temperature selection varies across the proteomes of these mussels. A possible explanation for this trend is that selection is most intense on classes of proteins that are especially perturbed by temperature or particularly important for maintaining homeostasis during thermal stress. The requirement for large conformational change during catalysis may render enzymes more susceptible to thermal stress than classes of protein where function is dependent on smaller movements [11, 19, 46]. Variation in hydrogen bonding and salt

bridge formation among *Mytilus* proteins provides support for this hypothesis. The set of *M. galloprovincialis* proteins with the greatest number of hydrogen bonds and salt bridges per amino acid contained a greater than expected proportion of enzymes. "Peptidase", "oxidoreductase", and "transferase", were the top over-represented molecular function ontologies among the 50% of *M. galloprovincialis* proteins with the most hydrogen bonds per amino acid. cMDH and IDH are enzymes that exhibit temperature-dependent differences in catalytic activity between *Mytilus* species consistent with temperature adaptation [45, 46]. Both cMDH and IDH are large, multi-subunit proteins that undergo substantial conformational shifts during ligand binding [5, 46], and both were among the *M. galloprovincialis* proteins with the most hydrogen bonds and salt bridges per amino acid (Additional file 1: Figures S1 and S2). Enzymatic bias in selection is also considered a causative factor for inconsistent temperature adaptation among the proteins of extremophiles. Proteome comparison of 24 prokaryotic species adapted to a wide range of temperatures showed that proteins annotating to "enzyme regulation" and "catalytic activity" ontologies experienced the greatest shifts in thermal stability [19].

Enzymes functioning within biochemical pathways that are crucial to surviving periods of thermal stress may also be subject to strong selection pressure. Antioxidant systems and programmed cell death are components of the conserved cellular stress response that is induced following temperature increase and that works to maintain homeostasis under suboptimal thermal conditions [78, 79]. Enzymes from these two important processes were over-represented among the small group of proteins with significantly different numbers of hydrogen bonds and salt bridges between *Mytilus* species. Over-represented ontologies relating to carboxylic acid metabolism were caused by high numbers of hydrogen bonds and salt bridges in cMDH and IDH of *M. galloprovincialis.* Although isoforms of these two enzymes participate in the tricarboxylic acid cycle (hence annotation to this ontology), the specific isoforms analyzed here function outside this pathway, and alternatively support the hypothesis that differences in the capacity of *M. galloprovincialis* and *M. trossulus* to cope with oxidative stress contributes to their divergent heat tolerances [52]. Exposure to heat stress increases the production of damaging reactive oxygen species (ROS) through an effect on the electron transport chain used for ATP production in the mitochondria [80]. Heat-induced increases in ROS can be ameliorated by reducing influx of NADH through the electron transport chain or by increasing production of antioxidants that can scavenge these molecules [74]. Both cMDH and IDH can influence these defense mechanisms: cMDH influences the influx of NADH to the mitochondria through the malate-aspartate shuttle and IDH catalyzes production of NADPH that is

needed by antioxidants like glutathione [74]. There is evidence to suggest that the higher heat tolerance of *M. galloprovincialis* compared with *M. trossulus* is due in part to an ability to control oxidative stress by enhancing the production of antioxidants rather than slowing the rate of NADH entry to the electron transport chain (and subsequently limiting ATP production) [52, 81]. Enhancing the thermal stability of enzymes like IDH and cMDH would allow chemical reactions that reduce oxidative stress and sustain ATP production to continue at higher temperatures in *M. galloprovincialis* compared with *M. trossulus*. This hypothesis is consistent with predicted increases in the total number of hydrogen bonds in *M. galloprovincialis* orthologs of IDH ($n = 1646$) and cMDH ($n = 1326$) compared with IDH ($n = 1644$) and cMDH ($n = 1320$) of *M. trossulus* (Additional file 1: Figures S1 and S2). Furthermore, other antioxidant enzymes were among the upper 25% of *M. galloprovincialis* proteins with the most hydrogen bonds and salt bridges per amino acid, including superoxide dismutase and glutathione S-transferase, suggesting that these enzymes may also be adapted to retain function at higher temperatures in *M. galloprovincialis* than in *M. trossulus*.

Programmed cell death (apoptosis) can be used to eliminate dysfunctional cells when macromolecular damage from thermal stress exceeds a cell's capacity for repair [78, 79]. Apoptosis is an important aspect of the response to heat stress in *Mytilus* mussels and differences in the abundance and activity of apoptotic proteins appear to contribute to interspecific differences in temperature tolerance [82]. Greater numbers of hydrogen bonds and salt bridges within apoptotic proteins could influence temperature tolerance of *Mytilus* mussels by allowing removal of irreparable cells to continue in *M. galloprovincialis* beyond temperatures that had caused proteins within cell death pathways to become non-functional in *M. trossulus*. *M. galloprovincialis* proteins that could exert a strong influence on cell fate during heat stress, including p63, DNA damage regulated protein (PDRP), and multiple caspase enzymes [72, 73, 83], had more hydrogen bonds and salt bridges per amino acid than *M. trossulus* proteins or other functional classes of *M. galloprovincialis* proteins. p63 is a member of the p53 family of transcription factors that serve as connection points for several stress response pathways and that govern the decision between cell survival and apoptosis by controlling the abundance of gene products that determine cell fate [70, 71, 73, 84, 85]. PDRP is a downstream target of p53, suggesting that this molecule also contributes to the balance between cell survival and death during environmental stress [73, 86]. Caspases are peptidases that degrade diverse substrates as part of apoptosis [72, 87, 88], and activity of these enzymes is necessary for removal of damaged macromolecules during heat stress. Caspases are also dependent on a large

conformational change to convert from an inactive to active form [89], which is consistent with structural changes associated with temperature adaptation occurring disproportionately in proteins with this characteristic.

Differences in amino acid use among *Mytilus* proteins

Prokaryotes living in extreme environments exhibit signatures of temperature adaptation in the amino acid composition of their proteins [6, 7, 9]. For proteins adapted to extreme heat, this imprint includes a higher proportion of hydrophobic, charged, and large volume residues, as well as an increase in the number of cysteines and prolines, and a decrease in glycine content [1, 3, 4]. Differences in amino acid use between *M. galloprovincialis* and *M. trossulus* indicate that some, but not all, of these adaptive mechanisms contribute to thermal adaptation in *Mytilus* proteins. Selection for protein function under considerably different temperature regimes is a plausible explanation for inconsistencies in amino acid use between *Mytilus* and prokaryotic extremophiles. Changes in protein stability that underlie small differences in habitat temperature and organism thermal tolerance, such as that occurring between *M. galloprovincialis* and *M. trossulus*, may favor particular mechanisms of structural change or use mechanisms observed in extremophiles much less frequently.

Increasing stability via addition of hydrogen bonds appears to be a method of high temperature adaptation used in proteins of both *Mytilus* mussels and thermophilic prokaryotes. Hydrogen bonds are found in greater abundance in thermophile proteins [32, 33], are less abundant in psychrophilic proteins [41–43], and consistent with this pattern, the upper 25 and 50% of *M. galloprovincialis* proteins with the most hydrogen bonds per amino acid had significantly more hydrogen bonds than did the upper 25 and 50% of *M. trossulus* proteins with the most hydrogen bonds per amino acid. Relative increases in hydrogen bonding among *M. galloprovincialis* proteins were accompanied by a net increase in amino acids most likely to form hydrogen bonds. Threonine, asparagine, and glutamine exhibited large proportional increases in *M. galloprovincialis* proteins relative to *M. trossulus* proteins, and all three of these residues have both donor and acceptor atoms in their side chains that facilitate formation of hydrogen bonds [90].

Increasing the number of salt bridges within proteins represents another mechanism of high temperature adaptation shared by *Mytilus* and thermophile proteins. The stabilizing effect of salt bridges has long been implicated in the evolution of heat tolerant proteins [91–94], and the upper 25 and 50% of *M. galloprovincialis* proteins had significantly more salt bridges per amino acid than did the upper 25 and 50% of *M. trossulus* proteins. Because salt bridges are the result of electrostatic interaction between

Chao *et al. BMC Evolutionary Biology* (2020) 20:28

Page 13 of 18

ionized groups of opposite charge [95], the relative increase in salt bridges among *M. galloprovincialis* proteins was expected to have occurred through substitution of charged for uncharged amino acids. A greater proportion of charged residues is one of the most frequently observed characteristics of thermophilic proteins [7, 13–17] and temperature-related differences in the function of IDH orthologs between *Mytilus* congeners were attributed to insertion of charged residues [46]. However, the indication from analysis of larger numbers of *Mytilus* proteins is that temperature adaptation in these species is not as simple as adding charged residues. Contrary to expectations, there was an overall decrease in the proportion of charged amino acids in the upper 25 and 50% of *M. galloprovincialis* proteins with the most salt bridges per amino acid relative to *M. trossulus*. Opposing directions of change between positively and negatively charged residues may explain this counterintuitive result. The overall decrease in charged amino acids among *M. galloprovincialis* proteins was principally caused by large declines in positively charged lysine and arginine; in contrast, the total number of negatively charged aspartate and glutamate increased in *M. galloprovincialis* proteins relative to *M. trossulus* proteins. Despite the decrease in lysine use in *M. galloprovincialis* relative to *M. trossulus*, lysine content remained high among the proteins of both species when compared with other amino acids: 6.7% of all residues in the upper 50% of *M. galloprovincialis* proteins with the most salt bridges per amino acid were lysine (fourth highest use rate) and 8.7% of residues in *M. trossulus* proteins with the most salt bridges per amino acid were lysine (highest use rate) (Table 5). Given that lysine outnumbered most other amino acids in these sets of *Mytilus* proteins, it is probable that most positively charged residues were not interacting with a negatively charged residue to form a salt bridge. Heat adaptation in *M. galloprovincialis* may have therefore involved an increase in the proportion of negatively charged amino acids so that new electrostatic interactions could form with the excess lysines already present, ultimately resulting in new salt bridges and more stable proteins compared with *M. trossulus*.

Mytilus congeners did not appear to modulate either hydrophobicity or volume in a majority of the proteins analyzed here, revealing a potential dichotomy between mechanisms of protein temperature adaptation used by extremophiles and those used by *Mytilus*. Proteins adapted to extreme heat have been repeatedly shown to increase the content of non-polar, hydrophobic amino acids at the expense of polar, hydrophilic residues such as serine, threonine, asparagine, and glutamine [6, 13, 18]. One analysis showed that thermophilic proteins preferred nearly every other amino acid to uncharged, polar residues, with removal of serine, threonine, asparagine, or glutamine accounting for 52 of 64 possible replacements in

the proteins analyzed [6]. This pattern of substitution resulted in a net loss of nine uncharged polar residues in a typical 300 amino acid thermophilic protein [6]. Contrary to these trends, the content of polar, more hydrophilic amino acids increased in *M. galloprovincialis* proteins compared with proteins of *M. trossulus*. The increase in polar amino acids is presumed to have occurred as a means of enhancing stability through additional hydrogen bonding, rather than increasing hydrophobicity, a premise that has support in *Mytilus*. The single amino acid substitution responsible for differences in the Km-temperature relationship of cMDH between *M. galloprovincialis* and *M. trossulus* was the result of a non-polar (valine) to polar (asparagine) substitution and formation of an additional hydrogen bond [45]. There was also significant positive correlation between the content of polar residues and organism temperature tolerance among cMDH orthologs of marine mollusks that included both *Mytilus* species [12].

Contrasting trends of amino acid use between *Mytilus* and extremophiles suggests that different mechanisms of evolution are at work when proteins adapt to moderate temperatures than when proteins adapt to temperature extremes [12]. Maximizing the hydrophobic effect may be required for the evolution of exceptionally stable proteins that function at the highest end of temperatures found in the biosphere. However, hydrophobicity may be less important when proteins adapt to more modest heat stress. In support of this conjecture, the hydrophobic effect makes a larger contribution to protein stability than does hydrogen bonding [96–98]. Furthermore, the upper thermal limit of protein function may not be the factor directly under selective pressure in *Mytilus* mussels, as the upper critical temperatures of these species is likely lower than the denaturation temperature of most of their proteins [5, 45, 46, 99]. Instead, selection may be acting to make small adjustments in local flexibility such that critical aspects of protein function, such as substrate binding or catalytic rate, are optimized at different temperatures [5, 46]. Changes in hydrogen bonding may be all that is required for small shifts in the optimal operating temperatures of *M. galloprovincialis* and *M. trossulus* proteins.

Shifts in the abundance of individual amino acids have also been correlated with increases in protein thermal stability and enhanced function at high temperatures [3, 4, 100]. Increasing the number of cysteines can enhance protein stability through the formation of disulfide bonds between sulfur atoms on adjacent residues [101, 102]. There were 109 more cysteine residues than expected in all *M. galloprovincialis* proteins, implying that additional disulfide bonds contribute to greater protein thermal stability and organism temperature tolerance in this species compared with *M. trossulus*. Disulfide bond formation generally occurs within the endoplasmic

Chao *et al. BMC Evolutionary Biology*   (2020) 20:28

Page 14 of 18

reticulum, and consequently, potentially adaptive shifts in disulfide bonding may be predominantly occurring in a smaller subset of extracellular *Mytilus* proteins [103]. Glycine content tends to decrease in thermophilic proteins because the absence of a sidechain in this residue imparts greater conformational flexibility that is assumed to be selected against at high temperatures [4, 100]. The increase in glycine content of *M. galloprovincialis* proteins compared with *M. trossulus* proteins (39 more residues than expected) suggests this adaptive mechanism does not affect the majority of proteins analyzed here. Glycine use in *Mytilus* more closely resembles that observed among mollusk orthologs of cMDH, where glycine content was positively correlated with organism temperature tolerance. In this case, addition of glycine to cMDH of heat-tolerant mollusks was theorized to organize the folded protein into a more compact and temperature-resistant structure [12]. Thermal stability of mollusk cMDH orthologs was also positively correlated with total proline content, which was predicted to slow heat denaturation by stabilizing sites of contact between subunits of the enzyme [12]. *M. galloprovincialis* proteins did not contain a higher than expected number of proline residues compared with *M. trossulus* proteins. Differences in proline content between *Mytilus* and other mollusk species may be another indication that certain types of adaptive change are favored at different magnitudes of temperature stress. The strongest evidence for the functional importance of proline in mollusk cMDH resulted from comparison of two snails (*Echinolittorina malaccana* and *Littorina keenae*) that have much higher temperature tolerances (LT$_{50}$ = 57 °C and LT$_{50}$ = 48 °C, respectively) than either *M. galloprovincialis* or *M. trossulus* (LT$_{50}$ = 38 °C and LT$_{50}$ = 34 °C, respectively) [12].

Caveats and future directions
Conclusions about the evolutionary influence of temperature on the structure of *Mytilus* proteins should be considered in the context of several caveats. A primary motivation for this study was to analyze potentially adaptive change over a very large number of proteins. The methodology used here was successful in collecting data from a sufficient number of proteins to provide insight into the proportion of the proteome modified during temperature adaptation, detect functional bias among proteins acted on by natural selection during temperature adaptation, and identify mechanisms of protein temperature adaption that are conserved or divergent across phylogeny. However, disparity in protein sequence resources between congeners, lack of homology to proteins with solved three-dimensional structures, and duplicate sequences in the UniProt database, reduced the number of *Mytilus* proteins that could be

analyzed. While a far greater number of proteins were evaluated in this study compared with most previous work, it is possible that the 58 *M. trossulus* proteins did not represent the full variation of hydrogen bonding, salt bridge formation, or amino acid use found across the proteome of this species. Differences in the number of hydrogen bonds and salt bridges between *Mytilus* proteins could therefore have resulted from omission of certain *M. trossulus* proteins rather than adaptive change in *M. galloprovincialis*. Sequencing the genomes of these species would increase the number of proteins that could be compared and allow this uncertainty to be empirically tested [104].

The methodology used here identified differences in the total number of hydrogen bonds, salt bridges and amino acids among *Mytilus* proteins, but did not include an evaluation of the location of adaptive change within proteins. Structural changes occurring in specific regions of *Mytilus* proteins could have greater impacts on thermal sensitivity than change occurring elsewhere in the protein. Amino acid substitutions that influence the stability of mobile regions, regions of subunit interaction, or coiled-coil domains, have all been hypothesized as important for temperature adaptation of *Mytilus* proteins [5, 12, 53]. Adaptive amino acid substitutions occurring among orthologs of dehydrogenase enzymes from marine organisms adapted to different temperatures tend to occur on the solvent exposed surface, rather than other regions of the protein [5]. Furthermore, adaptation through adjustments in local-scale protein stability can occur without a concomitant change in global protein stability [46, 99, 105–107], suggesting that it is possible for adaptive change to occur without altering the total number of hydrogen bonds or salt bridges or without obvious deviations in amino acid composition. Development of high-throughput approaches that can distinguish stabilizing interactions in specific regions of proteins (e.g. surface vs. core) is necessary to identify hotspots of evolutionary change.

The ability of in silico modeling algorithms to accurately capture subtle changes in protein structure has been questioned [5, 12, 108]. Single amino acid substitutions have been repeatedly shown to underlie structural and functional differences in the enzymes of marine ectotherms, including between *Mytilus* mussels [8, 45, 109, 110]. Orthologs of IDH differ by only two non-conservative amino acid substitutions between *M. galloprovincialis* and *M. trossulus*, and orthologs of cMDH differ by only a single non-conservative change between the two species [45, 46]. The methodology used here predicted a small increase in the total number of hydrogen bonds in *M. galloprovincialis* orthologs of IDH and cMDH compared with orthologs of *M. trossulus*, a result that is consistent with small differences in the amino acid sequences of these orthologs and

Chao *et al. BMC Evolutionary Biology*   (2020) 20:28

Page 15 of 18

the temperature tolerances of each species. However, comparison of structure-function relationships among other mollusk cMDH shows that in silico predictions of protein stability do not always replicate in vitro measures of thermal performance [5, 12]. High-throughput experiments that couple in silico with in vitro or in vivo analyses may address this problem, but will require development of new methods capable of assessing larger numbers of proteins in vitro or in vivo.

The presence or absence of ligands in *Mytilus* protein models could also influence patterns of hydrogen bonding described in this study. The HBOND algorithm used here considers the atomic distance of all potential donor and acceptor atoms listed in the .pdb file, including those present in the ligand [60]. A *Mytilus* protein model with the ligand included may therefore have a different number of hydrogen bonds compared with a model of the same protein without the ligand. The methodology used here prioritized selection of the most accurate template and did not account for the presence or absence of ligands in the chosen template. This approach was chosen to maximize the number of proteins that could be analyzed, rather than restrict analysis to a much smaller subset of proteins with or without ligands. Future analyses may allow for the effect of ligands on hydrogen bonding and salt bridge formation to be explicitly quantified if more protein sequences are obtained for *M. galloprovincialis* and *M. trossulus*.

A final consideration is that *Mytilus* species compared here also differ in their tolerance of salinity: *M. galloprovincialis* is more tolerant of increases in salinity and *M. trossulus* is more tolerant of decreases in salinity [44, 50]. Some of the protein structural changes examined here could therefore reflect selection pressure resulting from differences in habitat salinity rather than temperature. The proteins of salt-tolerant halophiles can be enriched with acidic amino acids, such as aspartate and glutamate, because side-chains from these negatively charged residues compete with cations in the environment and prevent protein aggregation through electrostatic repulsion [111]. The proportion of aspartate and glutamate was higher in more salt-tolerant *M. galloprovincialis* proteins than in less salt-tolerant *M. trossulus* proteins. Additional experiments are required to separate proteins contributing to thermal tolerance from those that underlie salinity tolerance in *Mytilus*.

Conclusions

Comparison of large numbers of proteins between warm-adapted *M. galloprovincialis* and cold-adapted *M. trossulus* indicates that protein structures from these two species vary in a manner that is consistent with adaptation to different temperatures, but that does not always conform to mechanisms of temperature adaptation observed in extremophiles. Only a portion of the proteins examined exhibited significant differences in the number of hydrogen bonds and salt bridges between *Mytilus* congeners, implying that comprehensive modification of the proteome is not required for small shifts in temperature tolerance and that adaptive responses to climate warming may be feasible in these species [112]. Enzymes, in particular those involved in oxidative stress responses and programmed cell death, were over-represented among proteins with the greatest number of hydrogen bonds and salt bridges in more heat tolerant *M. galloprovincialis*. This result is indicative of bias during temperature adaptation toward proteins in which function is dependent on large conformational change [46, 99, 107]. Structural changes in *Mytilus* proteins were associated with shifts in amino acid use that were likely to alter the number of hydrogen bonds, salt bridges, and disulfide bonds, but unlikely to substantially change hydrophobicity, volume, or proline and glycine content. Evolutionary changes required to maintain protein function in response to small shifts in environmental temperature are therefore likely to differ from those needed for function at temperature extremes.

Supplementary information

Additional file 1: Figure S1. All unique proteins QMEAN >-5 *M. galloprovincialis*. **Figure S2.** All unique proteins QMEAN >-5 *M. trossulus*. **Figure S3.** All unique orthologs QMEAN >-5. **Figure S4.** Over-representation of *M. galloprovincialis* protein ontologies. **Figure S5.** Spearman rank correlation between the number of hydrogen bonds and salt bridges per amino acid among *Mytilus* proteins. **Figure S6.** Amino acid composition all unique proteins *M. galloprovincialis*. **Figure S7.** Amino acid composition all unique proteins *M. trossulus*. **Figure S8.** Chi square tests of independence comparing amino acid use in different sets of *Mytilus* proteins. **Figure S9.** Spearman rank order correlation between residue hydrophobicity or residue volume and deviation in amino acid use in *M. galloprovincialis* relative to *M. trossulus* for different sets of proteins.

Abbreviations

ATP: Adenosine triphosphate; cMDH: Cytosolic malate dehydrogenase; DNA: Deoxyribonucleic acid; ESBRI: Evaluating Salt BRIdges in proteins; IDH: Isocitrate dehydrogenase; Km: Michaelis-Menten constant; LT_{50}: Lethal temperature for 50% of tested organisms; NADH: Nicotinamide adenine dinucleotide; NADPH: Nicotinamide adenine dinucleotide phosphate; pdb: Protein database; PDRP: P53 and DNA damage regulated protein; QMEAN: Qualitative Model Energy ANalysis; ROS: Reactive oxygen species

Chao *et al. BMC Evolutionary Biology*   (2020) 20:28

Page 16 of 18

the study, in collection, analysis, and interpretation of data, or in writing the manuscript.

References

1. Somero GN, Lockwood BL, Tomanek L. Biochemical adaptation: response to environmental challenges form life's origins to the Anthropocene. Sunderland: Sinauer Associates; 2017. Available from: https://global.oup.com/ukhe/product/biochemical-adaptation-9781605355641?cc=us&lang=en&
2. Somero GN. Proteins and temperature. Annu Rev Physiol. 1995;57(1):43–68.
3. Fields PA. Protein function at thermal extremes: balancing stability and flexibility. Comp Biochem Physiol A Mol Integrat Physiol. 2001;129(2–3):417–31.
4. Sælensminde G, Halskau Ø, Jonassen I. Amino acid contacts in proteins adapted to different temperatures: hydrophobic interactions and surface charges play a key role. Extremophiles. 2009;13(1):11.
5. Fields PA, Dong Y, Meng X, Somero GN. Adaptations of protein structure and function to temperature: there is more than one way to 'skin a cat'. J Exp Biol. 2015;218(12):1801–11.
6. Haney PJ, Badger JH, Buldak GL, Reich CI, Woese CR, Olsen GJ. Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. Proc Natl Acad Sci U S A. 1999; 96(7):3578–83.
7. Zeldovich KB, Berezovsky IN, Shakhnovich EI. Protein and DNA sequence determinants of thermophilic adaptation. PLoS Comp Biol. 2007;3(1):e5.
8. Dong Y, Somero GN. Temperature adaptation of cytosolic malate dehydrogenases of limpets (genus *Lottia*): differences in stability and function due to minor changes in sequence correlate with biogeographic and vertical distributions. J Exp Biol. 2009;212(2):169–77.
9. Ma BG, Goncearenco A, Berezovsky IN. Thermophilic adaptation of protein complexes inferred from proteomic homology modeling. Structure. 2010; 18(7):819–28.
10. Liao ML, Zhang S, Zhang GY, Chu YM, Somero GN, Dong YW. Heat-resistant cytosolic malate dehydrogenases (cMDHs) of thermophilic intertidal snails (genus *Echinolittorina*): protein underpinnings of tolerance to body temperatures reaching 55 C. J Exp Biol. 2017;220(11):2066–75.
11. Dong YW, Liao ML, Meng XL, Somero GN. Structural flexibility and protein adaptation to temperature: molecular dynamics analysis of malate dehydrogenases of marine molluscs. Proc Natl Acad Sci U S A. 2018;115(6): 1274–9.
12. Liao ML, Somero GN, Dong YW. Comparing mutagenesis and simulations as tools for identifying functionally important sequence changes for protein thermal adaptation. Proc Natl Acad Sci U S A. 2019;116(2):679–88.
13. Cambillau C, Claverie JM. Structural and genomic correlates of hyperthermostability. J Biol Chem. 2000;275(42):32383–6.
14. Das R, Gerstein M. The stability of thermophilic proteins: a study based on comprehensive genome comparison. Funct Integr Genomic. 2000;1(1):76–88.
15. Chakravarty S, Varadarajan R. Elucidation of factors responsible for enhanced thermal stability of proteins: a structural genomics based study. Biochemistry. 2002;41(25):8152–61.
16. Tekaia F, Yeramian E, Dujon B. Amino acid composition of genomes, lifestyles of organisms, and evolutionary trends: a global picture with correspondence analysis. Gene. 2002;297(1–2):51–60.
17. Singer GA, Hickey DA. Thermophilic prokaryotes have characteristic patterns of codon usage, amino acid composition and nucleotide content. Gene. 2003;317:39–47.
18. Berezovsky IN, Shakhnovich EI. Physics and evolution of thermophilic adaptation. Proc Natl Acad Sci U S A. 2005;102(36):12742–7.
19. Gu J, Hilser VJ. Sequence-based analysis of protein energy landscapes reveals nonuniform thermal adaptation within the proteome. Mol Biol Evol. 2009;26(10):2217–27.
20. IPCC. Summary for policymakers. In: Masson-Delmotte V, Zhai P, Pörtner H-O, Roberts D, Skea J, Shukla PR, Pirani A, Moufouma-Okia W, Péan C, Pidcock R, Connors S, Matthews JBR, Chen Y, Zhou X, Gomis MI, Lonnoy E, Maycock T, Tignor M, Waterfield T, editors. Global warming of 1.5°C, An IPCC special report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty; 2018. Available from: https://www.ipcc.ch/sr15/.
21. Goldenzweig A, Fleishman SJ. Principles of protein stability and their application in computational design. Annu Rev Biochem. 2018;87:105–29.
22. Zavodszky P, Kardos J, Svingor Á, Petsko GA. Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. Proc Natl Acad Sci U S A. 1998;95(13):7406–11.
23. Howell SC, Inampudi KK, Bean DP, Wilson CJ. Understanding thermal adaptation of enzymes through the multistate rational design and stability prediction of 100 adenylate kinases. Structure. 2014;22(2):218–29.
24. Sikosek T, Chan HS. Biophysics of protein evolution and evolutionary protein biophysics. J R Soc Interface. 2014;11(100):20140419.
25. D'Amico S, Marx JC, Gerday C, Feller G. Activity-stability relationships in extremophilic enzymes. J Biol Chem. 2003;278(10):7891–6.
26. Beadle BM, Shoichet BK. Structural bases of stability–function tradeoffs in enzymes. J Mol Biol. 2002;321(2):285–96.
27. Somero GN. Adaptation of enzymes to temperature: searching for basic "strategies". Comp Biochem Physiol B: Biochem Mol Biol. 2004;139(3):321–33.
28. Somero GN. The physiology of global change: linking patterns to mechanisms. Annu Rev Mar Sci. 2012;4:39–61.
29. Feller G, Arpigny JL, Narinx E, Gerday C. Molecular adaptations of enzymes from psychrophilic organisms. Comp Biochem Physiol A Physiol. 1997; 118(3):495–9.
30. D'Amico S, Gerday C, Feller G. Temperature adaptation of proteins: engineering mesophilic-like activity and stability in a cold-adapted α-amylase. J Mol Biol. 2003;332(5):981–8.
31. Feller G. Protein stability and enzyme activity at extreme biological temperatures. J Phys Condens Matter. 2010;22(32):323101.
32. Vogt G, Woell S, Argos P. Protein thermal stability, hydrogen bonds, and ion pairs. J Mol Biol. 1997;269(4):631–43.
33. Kumar S, Tsai CJ, Nussinov R. Factors enhancing protein thermostability. Protein Eng. 2000;13(3):179–91.
34. Berezovsky IN, Zeldovich KB, Shakhnovich EI. Positive and negative design in stability and thermal adaptation of natural proteins. PLoS Comp Biol. 2007;3(3):e52.
35. Van Dijk E, Hoogeveen A, Abeln S. The hydrophobic temperature dependence of amino acids directly calculated from protein structures. PLoS Comp Biol. 2015;11(5):e1004277.
36. Zanphorlin LM, De Giuseppe PO, Honorato RV, Tonoli CC, Fattori J, Crespim E, De Oliveira PS, Ruller R, Murakami MT. Oligomerization as a strategy for cold adaptation: structure and dynamics of the GH1 β-glucosidase from *Exiguobacterium antarcticum* B7. Sci Rep. 2016;6:23776.
37. Gianese G, Argos P, Pascarella S. Structural adaptation of enzymes to low temperatures. Protein Eng. 2001;14(3):141–8.
38. Kumar S, Nussinov R. Salt bridge stability in monomeric proteins. J Mol Biol. 1999;293(5):1241–55.
39. Kumar S, Tsai CJ, Ma B, Nussinov R. Contribution of salt bridges toward protein thermostability. J Biolmol Struct Dyn. 2000;17(sup1):79–85.
40. Pucci F, Rooman M. Improved insights into protein thermal stability: from the molecular to the structurome scale. Philos T R Soc A. 2016;374(2080):20160141.
41. Russell NJ. Toward a molecular understanding of cold activity of enzymes from psychrophiles. Extremophiles. 2000;4(2):83–90.
42. Feller G, Gerday C. Psychrophilic enzymes: hot topics in cold adaptation. Nat Rev Microbiol. 2003;1(3):200.
43. Siddiqui KS, Cavicchioli R. Cold-adapted enzymes. Annu Rev Biochem. 2006; 75:403–33.

44. Lockwood BL, Somero GN. Invasive and native blue mussels (genus *Mytilus*) on the California coast: the role of physiology in a biological invasion. J Exp Mar Biol Ecol. 2011;400(1–2):167–74.

45. Fields PA, Rudomin EL, Somero GN. Temperature sensitivities of cytosolic malate dehydrogenases from native and invasive species of marine mussels (genus *Mytilus*): sequence-function linkages and correlations with biogeographic distribution. J Exp Biol. 2006;209(4):656–67.

46. Lockwood BL, Somero GN. Functional determinants of temperature adaptation in enzymes of cold-versus warm-adapted mussels (Genus *Mytilus*). Mol Biol Evol. 2012;29(10):3061–70.

47. McDonald JH, Koehn RK. The mussels *Mytilus galloprovincialis* and *Mytilus trossulus* on the Pacific coast of North America. Mar Biol. 1988;99:111–8.

48. Vermeij GJ. Anatomy of an invasion: the trans-Arctic interchange. Paleobiology. 1991;17(3):281–307.

49. Seed R. Systematics evolution and distribution of mussels belonging to the genus *Mytilus*-an overview. Am Malacol Bull. 1992;9(2):123–37.

50. Braby CE, Somero GN. Following the heart: temperature and salinity effects on heart rate in native and invasive species of blue mussels (genus *Mytilus*). J Exp Biol. 2006;209(13):2554–66.

51. Lockwood BL, Sanders JG, Somero GN. Transcriptomic responses to heat stress in invasive and native blue mussels (genus *Mytilus*): molecular correlates of invasive success. J Exp Biol. 2010;213(20):3548–58.

52. Tomanek L, Zuzow MJ. The proteomic response of the mussel congeners *Mytilus galloprovincialis* and *M. trossulus* to acute heat stress: implications for thermal tolerance limits and metabolic costs of thermal stress. J Exp Biol. 2010;213(20):3559–74.

53. Saarman NP, Kober KM, Simison WB, Pogson GH. Sequence-based analysis of thermal adaptation and protein energy landscapes in an invasive blue mussel (*Mytilus galloprovincialis*). Genome Biol Evol. 2017;9(10):2739–51.

54. Somero GN. Comparative physiology: a "crystal ball" for predicting consequences of global change. Am J Physiol Regul Integr Comp Physiol. 2010;301(1):R1–4.

55. Fabiola F, Bertram R, Korostelev A, Chapman MS. An improved hydrogen bond potential: impact on medium resolution protein structures. Protein Sci. 2002;11(6):1415–23.

56. The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017;45(D1):D158–69.

57. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics. 2006;22(2):195–201.

58. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, Schwede T. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 2014;42(W1):W252–8.

59. Berman H, Henrick K, Nakamura H, Markley JL. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. Nucleic Acids Res. 2006;35(suppl_1):D301–3.

60. Mizuguchi K, Deane CM, Blundell TL, Johnson MS, Overington JP. JOY: protein sequence-structure representation and analysis. Bioinformatics. 1998; 14(7):617–23.

61. Costantini S, Colonna G, Facchiano AM. ESBRI: a web server for evaluating salt bridges in proteins. Bioinformation. 2008;3(3):137.

62. Benkert P, Tosatto SC, Schomburg D. QMEAN: a comprehensive scoring function for model quality assessment. Proteins. 2008;71(1):261–77.

63. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TA, Rempfer C, Bordoli L, Lepore R. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018; 46(W1):W296–303.

64. Gene Ontology Consortium. The gene ontology resource: 20 years and still GOing strong. Nucleic Acids Res. 2018;47(D1):D330–8.

65. Alexa A, Rahnenführer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics. 2006;22(13):1600–7.

66. Alexa A, Rahnenfuhrer J. topGO: enrichment analysis for gene ontology. R package version. 2019. 2.36.0.

67. Tsuji J, Nydza R, Wolcott E, Mannor E, Moran B, Hesson G, Arvidson T, Howe K, Hayes R, Ramirez M, Way M. The frequencies of amino acids encoded by genomes that utilize standard and nonstandard genetic codes. Bios. 2010;1: 22–31.

68. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982;157(1):105–32.

69. Zamyatnin A. Protein volume in solution. Prog Biophys Mol Biol. 1972;24: 107–23.

70. Cox RL, Stephens RE, Reinisch CL. p63/73 homologues in surf clam: novel signaling motifs and implications for control of expression. Gene. 2003;320:49–58.

71. Štifanić M, Mičić M, Ramšak A, Blašković S, Ruso A, Zahn RK, Batel R. p63 in *Mytilus galloprovincialis* and p53 family members in the phylum Mollusca. Comp Biochem Physiol B Biochem Mol Biol. 2009;154(3):264–73.

72. Romero A, Estevez-Calvar N, Dios S, Figueras A, Novoa B. New insights into the apoptotic process in mollusks: characterization of caspase genes in *Mytilus galloprovincialis*. PLoS One. 2011;6(2):e17003.

73. Estevez-Calvar N, Romero A, Figueras A, Novoa B. Genes of the mitochondrial apoptotic pathway in *Mytilus galloprovincialis*. PLoS One. 2013;8(4):e61502.

74. Tomanek L. Proteomic responses to environmentally induced oxidative stress. J Exp Biol. 2015;218(12):1867–79.

75. Scheiner S, Kar T, Pattanayak J. Comparison of various types of hydrogen bonds involving aromatic amino acids. J Am Chem Soc. 2002;124(44): 13257–64.

76. Brininger C, Spradlin S, Cobani L, Evilia C. The more adaptive to change, the more likely you are to survive: protein adaptation in extremophiles. Semin Cell Dev Biol. 2018;84:158–69.

77. Hickey DA, Singer GAC. Genomic and proteomic adaptations to growth at high temperature. Genome Biol. 2004;5(10):117.

78. Kültz D. Evolution of the cellular stress proteome: from monophyletic origin to ubiquitous function. J Exp Biol. 2003;206(18):3119–24.

79. Kültz D. Molecular and evolutionary basis of the cellular stress response. Annu Rev Physiol. 2005;67:225–57.

80. Abele D, Heise K, Pörtner HO, Puntarulo S. Temperature-dependence of mitochondrial function and production of reactive oxygen species in the intertidal mud clam *Mya arenaria*. J Exp Biol. 2002;205:1831–41.

81. Tomanek L. Environmental proteomics of the mussel *Mytilus*: implications for tolerance to stress and change in limits of biogeographic ranges in response to climate change. Integrat Comp Biol. 2012;52(5):648–64.

82. Yao CL, Somero GN. The impact of acute temperature stress on hemocytes of invasive and native mussels (*Mytilus galloprovincialis* and *Mytilus californianus*): DNA damage, membrane integrity, apoptosis and signaling pathways. J Exp Biol. 2012;215(24):4267–77.

83. Romero A, Novoa B, Figueras A. The complexity of apoptotic cell death in mollusks: an update. Fish Shellfish Immunol. 2015;46(1):79–87.

84. Muttray AF, Schulte PM, Baldwin SA. Invertebrate p53-like mRNA isoforms are differentially expressed in mussel haemic neoplasia. Mar Environ Res. 2008;66(4):412–21.

85. Roos WP, Thomas AD, Kaina B. DNA damage and the balance between survival and death in cancer biology. Nat Rev Cancer. 2016;16(1):20.

86. Luo X, Huang Y, Sheikh MS. Cloning and characterization of a novel gene PDRG that is differentially regulated by p53 and ultraviolet radiation. Oncogene. 2003;22(46):7247.

87. Crawford ED, Wells JA. Caspase substrates and cellular remodeling. Annu Rev Biochem. 2011;80:1055–87.

88. Green DR, Llambi F. Cell death signaling. Cold Spring Harb Perspect Biol. 2015;7(12):a006080.

89. Shi Y. Caspase activation, inhibition, and reactivation: a mechanistic view. Protein Sci. 2004;13(8):1979–87.

90. Kim K, Friesner RA. Hydrogen bonding between amino acid backbone and side chain analogues: a high-level ab initio study. J Am Chem Soc. 1997; 119(52):12952–61.

91. Elcock AH. The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. J Mol Biol. 1998;284(2):489–502.

92. Georlette D, Damien B, Blaise V, Depiereux E, Uversky VN, Gerday C, Feller G. Structural and functional adaptations to extreme temperatures in psychrophilic, mesophilic, and thermophilic DNA ligases. J Biol Chem. 2003; 278(39):37015–23.

93. Zhou HX, Dong F. Electrostatic contributions to the stability of a thermophilic cold shock protein. Biophys J. 2003;84(4):2216–22.

94. Thomas AS, Elcock AH. Molecular simulations suggest protein salt bridges are uniquely suited to life at high temperatures. J Am Chem Soc. 2004; 126(7):2208–14.

95. Barlow DJ, Thornton JM. Ion-pairs in proteins. J Mol Biol. 1983;168(4):867–85.

96. Baldwin RL. Energetics of protein folding. J Mol Biol. 2007;371(2):283–301.

97. Fitzpatrick AW, Knowles TP, Waudby CA, Vendruscolo M, Dobson CM. Inversion of the balance between hydrophobic and hydrogen bonding

interactions in protein folding and aggregation. PLoS Comp Biol. 2011;7(10): e1002169.

98. Pace CN, Fu H, Fryar KL, Landua J, Trevino SR, Shirley BA, Hendricks MM, Iimura S, Gajiwala K, Scholtz JM, Grimsley GR. Contribution of hydrophobic interactions to protein stability. J Mol Biol. 2011;408(3):514–28.

99. Fields PA, Somero GN. Hot spots in cold adaptation: localized increases in conformational flexibility in lactate dehydrogenase A4 orthologs of Antarctic notothenioid fishes. Proc Natl Acad Sci U S A. 1998;95(19):11476–81.

100. Sælensminde G, Halskau Ø, Helland R, Willassen NP, Jonassen I. Structure-dependent relationships between growth temperature of prokaryotes and the amino acid frequency in their proteins. Extremophiles. 2007;11(4):585–96.

101. Rosato V, Pucello N, Giuliano G. Evidence for cysteine clustering in thermophilic proteomes. Trends Genet. 2002;18(6):278–81.

102. Beeby M, O'Connor BD, Ryttersgaard C, Boutz DR, Perry LJ, Yeates TO. The genomics of disulfide bonding and protein stabilization in thermophiles. PLoS Biol. 2005;3(9):e309.

103. Bošnjak I, Bojović V, Šegvić-Bubić T, Bielen A. Occurrence of protein disulfide bonds in different domains of life: a comparison of proteins from the Protein Data Bank. Protein Eng Des Sel. 2014;27(3):65–72.

104. Murgarella M, Puiu D, Novoa B, Figueras A, Posada D, Canchaya C. A first insight into the genome of the filter-feeder mussel *Mytilus galloprovincialis*. PLoS One. 2016;11(3):e0151561.

105. Arnold FH, Wintrode PL, Miyazaki K, Gershenson A. How enzymes adapt: lessons from directed evolution. Trends Biochem Sci. 2001;26(2):100–6.

106. Daniel RM, Dunn RV, Finney JL, Smith JC. The role of dynamics in enzyme activity. Annu Rev Biophys Biomol Struct. 2003;32(1):69–92.

107. Bae E, Phillips GN. Structures and analysis of highly homologous psychrophilic, mesophilic, and thermophilic adenylate kinases. J Biol Chem. 2004;279(27):28202–8.

108. Bottaro S, Lindorff-Larsen K. Biophysical experiments and biomolecular simulations: a perfect match? Science. 2018;361(6400):355–60.

109. Holland LZ, McFall-Ngai M, Somero GN. Evolution of lactate dehydrogenase-a homologs of barracuda fishes (genus *Sphyraena*) from different thermal environments: differences in kinetic properties and thermal stability are due to amino acid substitutions outside the active site. Biochemistry. 1997; 36(11):3207–15.

110. Johns GC, Somero GN. Evolutionary convergence in adaptation of proteins to temperature: A4-lactate dehydrogenases of Pacific damselfishes (*Chromis* spp.). Mol Biol Evol. 2004;21(2):314–20.

111. Paul S, Bag SK, Das S, Harvill ET, Dutta C. Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes. Genome Biol. 2008;9(4):R70.

112. Palumbi SR, Evans TG, Pespeni MH, Somero GN. Present and future adaptation of marine assemblages. Oceanography. 2019;32(3):82–93.

Publisher's Note