Title Page

Classification: Biological Sciences (Physiology)

Title: Comparing mutagenesis and simulations as tools for identifying functionally important sequence changes for protein thermal adaptation

Short Title: Thermal adaptation of enzymes

Author Affiliations:

Ming-ling Liao[a], George N. Somero[b], and Yun-wei Dong[a,1]

[1]State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Sciences, Xiamen University, 361102 Xiamen, China; [2]Hopkins Marine Station, Department of Biology, Stanford University, Pacific Grove, CA 93950, USA.

Author contributions: M.L.L., G.N.S., and Y.W.D. designed the research, performed the experiments, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

Corresponding Author: [1]To whom correspondence should be addressed: Yun-wei Dong, Xiamen University, 361102 Xiamen, China. Tel: 86-592-2181680. Email: dongyw@xmu.edu.cn.

Keywords: adaptation | cytosolic malate dehydrogenase| evolution | molecular dynamics simulations | protein evolution

Abstract:

Comparative studies of orthologous proteins of species evolved at different temperatures have revealed consistent patterns of temperature-related variation in thermal stabilities of structure and function. However, the precise mechanisms by which interspecific variations in sequence foster these adaptive changes remain largely unknown. Here, we compare orthologs of cytosolic malate dehydrogenase (cMDH) from marine molluscs adapted to temperatures ranging from -1.9 °C (Antarctica) to ~55 °C (South China coast) and show how amino acid usage in different regions of the enzyme (surface, intermediate depth, and protein core) varies with adaptation temperature. This eukaryotic enzyme follows some, but not all of the rules established in comparisons of archaeal and bacterial proteins. To link the effects of specific amino acid substitutions with adaptive variations in enzyme thermal stability we combined site-directed mutagenesis (SDM) and *in vitro* protein experimentation with *in silico* mutagenesis using molecular dynamics simulation (MDS) techniques. SDM and MDS methods generally, but not invariably yielded common effects on protein stability. MDS analysis is shown to provide insights into how specific amino acid substitutions affect the conformational flexibilities of mobile regions (MRs) of the enzyme that are essential for binding and catalysis. Whereas these substitutions invariably lie outside of the MRs, they effectively transmit their flexibility-modulating effects to the MRs through linked interactions among surface residues. This discovery illustrates that regions of the protein surface lying outside of the site of catalysis can help establish an enzyme's thermal responses and foster evolutionary adaptation of function.

Significance Statement

Comparison of 26 cytosolic malate dehydrogenase (cMDH) orthologs of marine molluscs adapted to temperatures ranging from -1.9 °C (Antarctica) to ~55 °C (South China coast) shows how amino acid usage in different regions of the enzyme varies with adaptation temperature. *In vitro* site-directed mutagenesis approaches and *in silico* molecular dynamics simulations were compared as tools for deducing functionally important sequence changes. Whereas these key amino acid substitutions invariably lie outside of the mobile regions (MRs) essential for function, they transmit their flexibility-modulating effects to the MRs through linked interactions among surface residues. Thus regions of the protein surface lying outside of the site of catalysis can help establish an enzyme's thermal responses and foster evolutionary adaptation of function.

Protein structure and function are highly sensitive to changes in temperature. Because of these thermal sensitivities, temperature has had a major influence on protein evolution: orthologs of proteins from warm-adapted species commonly show different thermal responses from orthologs of cold-adapted species (1). For example, in a recent study, we compared the thermal stabilities of structure and cofactor binding ability in orthologs of cytosolic malate dehydrogenase (cMDH) from marine molluscs adapted to temperatures ranging from ~0 °C to ~55 °C (2). Consistent temperature-correlated trends were observed in all measured variables: rate of heat denaturation, thermal stability of the Michaelis-Menten constant (K_M) of cofactor (NADH) and structural flexibility as indexed by *in silico* molecular dynamics simulation (MDS) methods. Here, in an effort to link these differences with interspecific variation in protein sequence, we extend this analysis by comparing deduced amino acid sequences of 26 cMDH orthologs from 10 genera of marine molluscs adapted to an almost 60 °C range of temperatures: from -1.9 °C in McMurdo Sound, Antarctica, to ~55 °C in the rocky intertidal zone along the South China Sea (Table 1).

Our study addresses the following questions. First, at the most general level of analysis, we asked how the overall amino acid compositions (measured as percentages of different classes of amino acids) and the compositions of different regions of the protein (surface, intermediate depth, and core) differ in relation to adaptation temperature. Second, to develop structure-function linkages, we asked what sites in the proteins are strongly conserved in all species and what sites are highly variable and, therefore, potential sites of adaptive change? Third, to test conjectures about the potential involvement of a given amino acid substitution in adaptation to temperature, we performed two types of mutagenesis experiments. One involved conventional site-directed mutagenesis (SDM) followed by enzyme isolation and *in vitro* characterization of protein stability (rate of heat denaturation) and function (K_M^{NADH}). The second mutagenesis technique involved MDS approaches (3). We used this *in silico* method to study the effects of the same substitution used in SDM on the overall structural flexibility (rrmsd in backbone atom position) and residue-specific flexibility [rrms fluctuation (rmsf)] of the protein. We were especially interested in determining whether the thermal responses of the mutated protein seen *in vitro* (SDM approach) were modified in a similar way for the protein mutated *in silico.* If amino acid substitutions generated by SDM and MDS consistently give similar results, then a possibility opens up for using this *in silico* approach to test the importance of sequence differences in a simpler manner than the more laborious approach using SDM following by enzyme isolation and *in vitro* characterization.

The protein we used as our study system, the cytosolic paralog of malate dehydrogenase (E.C. 1.1.1.37), is a well-characterized protein for which many sequences and three-dimensional structures are available (4–7). The abundant data on cMDH allow one to study sites within the sequence that are known to be important in ligand binding, catalysis, and subunit-subunit interactions. The conformational changes essential for catalytic function are also well-understood, which facilitates structure-function analyses (6). Moreover, our recent work has identified regions of the protein that are key for establishing the appropriate levels of conformational flexibility, to allow the mobile regions (MRs) of the

protein to undergo the changes in conformation that are essential for ligand binding and catalysis (2, 3).

In this paper we first provide sequence information on the 26 cMDH orthologs and illustrate several trends between adaptation temperature and amino acid compositions of different regions of the protein. We then use these sequence data to develop a mechanistic analysis of structure-function relationships using SDM and MDS methods. We focus particular attention on four cMDH orthologs from two genera of littorinid snails, *Echinolittorina* and *Littorina*, whose cMDHs differ in thermal responses but have only minor differences in sequence. Through analyses of these four orthologs we show that single changes in amino acid sequence can affect thermal stability and ligand binding, but that an adaptive change in thermal stability is not necessarily linked to an adaptive change in binding and vice versa. We also show that there is a type of asymmetry in amino acid substitutions between heat-stable and heat-labile orthologs. Thus, for example, replacing a residue in a heat-stable protein with the residue found in a heat-labile protein may not weaken the structure, whereas the opposite replacement may significantly enhance stability. Comparing results of SDM and MDS analyses, we show that the effects obtained through *in vitro* studies with SDM-generated mutant proteins are generally, but not invariably reflected in the thermal responses shown *in silico* by MDS methods. However, MDS methods are shown to provide insights into the sites in the sequence where amino acid substitutions can lead to changes in flexibility of the mobile regions whose movements are essential for catalysis. These sites lie outside of the MR regions yet modulate their movements during function. The MDS studies thus provide evidence for involvement of many regions of the protein surface outside of the active site in modulating the stability and function of an enzyme and fostering temperature-adaptive evolutionary change.

RESULTS

Deduced Protein Sequences and Temperature-Related Differences in Amino Acid Compositions. In these 26 cMDH orthologs some functionally important regions, notably in residues known to interact with cofactor and substrates, and the active site of proton acceptor function, are completely conserved (Fig. 1, SI Appendix, Table S1). All of the "universally conserved residues" mentioned by Chapman et al. (6) (G14, I16, R92, R98, R161, H187, and S240 in the present study) are found in all of these molluscan orthologs. The residues involved in subunit-subunit interaction are also almost fully conserved; the only variable site in the region of subunit-subunit interaction is at site 22 (tyrosine, histidine or leucine). High variation in sequence occurs at many other sites including those conjectured (see below) to influence the flexibility of the MRs that undergo large changes in conformation during function (e.g. the highly variable site 114).

All 26 sequences were used in an analysis of overall- and region-specific variation in amino acid composition as a function of enzyme stability, as indexed by rate of loss of activity during incubation at elevated temperature (Fig. 2). As expected, protein thermal stability was strongly correlated with the species' adaptation temperatures (Fig. 3, R^2 = 0.720 and 0.866 for linear and quadratic regressions, respectively). The fractions of different categories of amino acids—charged residues, weakly hydrophobic/polar residues, and hydrophobic

residues—on the protein surface, at intermediate depth in the structure, and in the deep core region were calculated (Fig. 2 *A*–*T*). In the entire protein, there are significant decreases of charged residues ($P < 0.001$) and increases in weakly hydrophobic/polar residues ($P < 0.001$) with increasing protein thermal stability (Fig. 2 *A, B*). In the surface region, there are significant decreases of charged residues ($P = 0.001$) with rising thermal stability (Fig. 2 *D*). Total proline content of the 26 orthologs and proline abundance on the protein surface are significantly positively correlated with protein thermal stability ($P = 0.030$ and 0.029, respectively) (Fig. 2 *M, N*). For glycine, total content and surface content increase with thermal stability ($P = 0.028$ and 0.019, respectively) (Fig. 2 *Q, R*). The amino acid compositions of the intermediate and core regions did not exhibit temperature-related trends.

Three-dimensional models of a single cMDH subunit were constructed for 6 of the 26 species to elucidate where potentially significant substitutions occurred in the folded enzyme structure and to obtain insights into how these substitutions might affect the stability and function of the enzyme (Fig. 4, SI Appendix, Fig. S1). Of particular interest were substitutions that might affect the movements of the two MRs (structures enclosed by dashed lines in *E. malaccana* model) during function. Based on our previous MDS results (2), there are substantial differences among species and genera in the amount of side-chain movement in the two MRs and in how side-chain movement in the MRs responds to an increase in simulation temperature. Furthermore, a common pattern noted in comparisons of congeneric orthologs is a stronger effect of increase in temperature on one or both of the MRs in the more cold-adapted species. Here, we employed both *in silico* MDS and SDM-*in vitro* methods to test conjectures about the possible roles of specific differences in sequence in establishing such adaptive variation in thermal sensitivity.

Dimeric assemblies of cMDHs also were constructed to optimize conditions for this structure-function analysis (Fig. 5). The residues involved in catalytic and binding functions (Fig. 1) are mostly located near or within the two highly conserved MRs (Fig. 5 *A*). A principal focus of our analysis, then, was to interpret how substitutions might influence the conformational mobilities of the MRs. To this end, we focused on the orthologs from congeners of two genera of littorinid snails, *E. malaccana*, *E. radiata*, *L. keenae*, and *L. scutulata* (Fig. 5 *B*). There are only seven non-conservative amino acid substitutions among these four orthologs (residues 4, 33, 41, 48, 114, 219, and 321), and this high degree of sequence similarity makes these four orthologs a tractable study system to examine the importance of specific amino acid substitutions on the enzymes' thermal sensitivities. The non-conservative substitutions among *Echinolittorina* and *Littorina* cMDHs mainly occur near (but not within) the MR sites in the individual subunit or in regions of the subunit that, in the dimer, lie near the other subunit's MRs. Thus, these variable sites could be important in affecting the structural stability of both monomer and dimer. Intergeneric differences between *E. malaccana* and *L. keenae* cMDHs are only found at sites 4 and 114. Residue 4 is close to the residues involved in subunit-subunit interactions, where a proline occurs in the *Echinolittorina* cMDHs and alanine is found in the two *Littorina* orthologs (Fig. 1, 5). For the *Chlorostoma* orthologs, where interspecific differences in sequence are greater than for the

other two genera, the sites of non-conservative sequence differences again are found outside of—yet in close proximity to—the mobile regions (Fig. 4).

***In silico* Analysis: Global- and Localized-Flexibility of cMDH Orthologs in Wild Type and Mutant Enzymes.** To link changes in sequence with temperature-adaptive alterations in stability and function, we first used MDS methods to analyze the overall flexibility of the backbone atoms of selected orthologs (rmsd analysis) and the movements of each individual side-chain in the sequence (RMSF analysis). The predicted stabilities obtained from these *in silico* analyses were then compared to results of subsequent *in vitro* studies using variants obtained by SDM methods.

To investigate the roles of site 4 and site 114 in establishing the structural stability of the cMDH orthologs, we performed rmsd and RMSF analyses of three wild type cMDHs [*E. malaccana* (wt-*Em*), *E. radiata* (wt-*Er*), and *L. keenae* (wt-*Lk*)], and four mutant cMDHs (mut-G114S: glycine mutated to serine at site 114 for the *E. malaccana* ortholog; mut-P4A: proline mutated to alanine at site 4 for the *E. malaccana* ortholog; mut-S114G: serine mutated to glycine for the *E. radiata* ortholog, and mut-A4P: alanine mutated to proline for the *L. keenae* ortholog) (SI Appendix, Fig. S2, S3, Table S2, S3). There was no significant difference in rmsd values of backbone atom movement between mut-G114S and the wild type ortholog of *E. malaccana* at 57 °C (Fig. 6 *A*). However, there were significantly different ($P < 0.05$) rmsd values of backbone atom movements in other cases: rmsd values of mut-S114G were higher compared to wt-*Er* (Fig. 6 *B*–*C*), wt-*Em* (Fig. 6 *A*–*C*), and mut-G114S (Fig. 6 *A*–*C*) orthologs at a simulation temperature of 57 °C. The equilibrium (10–20 ns) rmsd value of mut-S114G was approximately equal to 2.7 Å, compared with 2.1 Å for wt-*Em*, 2.0 Å for mut-G114S, and 2.2 Å for wt-*Er* orthologs, indicating that the mutation from serine to glycine at site 114 significantly decreased thermal stability of the *E. radiata* ortholog. Replacing the alanine at site 4 with proline for the *L. keenae* ortholog (mut-A4P) significantly increased thermal stability, as indicated by the decrease in rmsd value of backbone atoms (Fig. 6 *D*–*F*). However, changing the proline at position 4 to an alanine in the *E. malaccana* ortholog (mut-P4A) did not change the rmsd value (Fig. 6 *D*–*F*).

RMSF results showed that *in silico* mutations had divergent effects on the residue-specific structural flexibility at a simulation temperature of 57 °C (Fig. 7 *A*–*D*). Compared to the wild type *E. malaccana* cMDH (wt-*Em*), mut-G114S exhibited no major changes in structural flexibility across the sequence; the two spectra were essentially superimposed (Fig. 7 *A*). However, compared to the *E. radiata* wild type (wt-*Er*), mut-S114G led to higher RMSF values around residues 45–55, indicating increased flexibility in this region as a consequence of replacing the serine with a glycine residue (Fig. 7 *B*). The two spectra were otherwise very similar, notably in the two MRs. The proline to alanine mutation at site 4 (mut-P4A) induced an increased flexibility around residues 20–35 compared to wt-*Em* (Fig. 6 *C*), a slight change in the specific residues showing maximal flexibility in MR1 (residues 90–105), and higher flexibility in residues 214–222, which lie just outside of MR2 (residues 230–245). In contrast to these effects seen with mut-P4A, when alanine is replaced by proline at site 4 for the *L. keenae* ortholog, lower rmsf values occur around residues 45–55 (Fig. 7 *D*), indicating a less flexible structure in this region.

***In vitro* analysis*: K_M^{NADH}* and Heat Denaturation of cMDH Orthologs in Recombinant Wild Type and Mutant Enzymes.** To determine if the MDS results reflect the properties of the enzymes measured *in vitro*, three recombinant wild types [*E. malaccana* (wt-*Em*), *E. radiata* (wt-*Er*), and *L. keenae* (wt-*Lk*)], and four mutant enzymes [mut-G114S and mut-P4A (corresponding to *E. malaccana* cMDH), mut-S114G (corresponding to the *E. radiata* ortholog), mut-A4P (corresponding to *L. keenae* ortholog)] were constructed for measurements of K_M^{NADH} and thermal stability (See SI Appendix, Table S4 for statistical results). Significant effects on thermal stability of cofactor (NADH) binding were revealed for some of the mutant enzymes. The K_M^{NADH} value of mut-G114S was significantly higher than that of wt-*Em* at 25°, 30°, 35°, and 40°C (P = 0.033, 0.029, < 0.001 and < 0.001, respectively) (Fig. 8 *A*), indicative of a stronger perturbation by temperature of the mut-G114S enzyme. Similarly, compared to wt-*Er*, mut-S114G had higher K_M^{NADH} values at measurement temperatures of 30 °C and 35 °C (P = 0.049 and 0.010, respectively). Across the full range of experimental temperatures, the slope of the relationship between K_M^{NADH} and measurement temperature for wt-*Em* was significantly lower than those of mut-G114S, wt-*Er*, and mut-S114G, further indicating that the wild type *E. malaccana* ortholog has the highest thermal stability in cofactor binding (P = 0.017, 0.002, and < 0.001, respectively) (Fig. 8 *A*, inset).

Mutations that exchanged proline and alanine residues also revealed significant effects on kinetics: The K_M^{NADH} value of mut-P4A cMDH was significantly higher than that of wt-*Em* at all temperatures (P < 0.05), indicating that mutagenesis from proline to alanine at position 4 for the *E. malaccana* ortholog decreased ligand binding ability (Fig. 8 *B*). A significant difference in the slope of the K_M^{NADH} versus temperature relationship also was observed between wt-*Em* and mut-P4A orthologs (P = 0.018) (Fig. 8 *B*, inset). In contrast to the effects observed with the mut-P4A substitution, the K_M^{NADH} value of mut-A4P showed no significant difference from that of wt-*Lk*, indicating this mutation doesn't change ligand binding ability (Fig. 8 *B*).

Thermal denaturation at 57.5 °C of recombinants wt-*Em*, wt-*Er*, wt-*Lk*, and the mutants mut-G114S, mut-S114G, mut-P4A, and mut-A4P revealed the relative importance of different residues at sites 4 and 114 on thermal stability (Fig. 9). Mutagenesis from glycine to serine at position 114 (mut-G114S) for the *E. malaccana* ortholog didn't change the rate of loss of activity (Fig. 9 *A*). In contrast, changing serine to glycine in the *E. radiata* enzyme (mut-S114G) led to a significant decrease in thermal stability (Fig. 9 *A*). The slope of the regression of residual activity as a function of incubation time of mut-S114G was significantly steeper than the slopes of wt-*Em*, wt-*Er*, and mut-G114S orthologs, indicating that the substitution of serine for glycine in the *E. radiata* ortholog can dramatically reduce the enzyme's thermal stability (P < 0.001, P < 0.001, and P = 0.002, respectively) (Fig. 9 *A*, inset).

Proline at site 4 is very important for the thermal stability of *E. malaccana* and *L. keenae* cMDHs (Fig. 9 *B*). When this proline is replaced by an alanine for *E. malaccana*, the thermal stability significantly decreased at 57.5 °C (wt-*Em* vs. mut-P4A) (P < 0.001) (Fig. 9 *B* inset). For *L. keenae*, thermal stability of the mutated ortholog in which alanine is replaced by a proline (mut-A4P) is significantly increased compared to the wild type (wt-*Lk*) (P = 0.031) (Fig. 9 *B*).

Discussion

In general, warm-adapted proteins are characterized by an enhanced global conformational rigidity, whereas cold-adapted proteins usually exhibit a relatively high flexibility (2, 8, 9). Previous studies that have sought to identify broad trends in temperature-related variation in amino acid composition and sequence that could explain these trends have focused primarily on proteins of archaeal and bacterial species whose optimal growth temperatures (OGT) spanned a range of over 100 °C (10). To our knowledge, no previous study of eukaryotic proteins has sampled orthologs from organisms with the wide range of adaptation temperatures, almost 60 °C, found in the present study of marine molluscs. Thus, our data allow generalizations about temperature-related amino acid compositions of orthologs of an enzyme from eukaryotes with body temperatures that span almost the full range of eukaryotic body temperatures (1) and also permit comparisons between trends found in eukaryotic and bacterial-archaeal proteins. Our data also provide a platform for designing mutagenesis studies to test the importance of specific substitutions on structural and functional thermal stability.

At the descriptive level, our data reveal patterns of temperature-related differences in fractional contributions of different classes of amino acids in specific regions of the protein and identify temperature-related changes in two amino acids, proline and glycine, that are likely to be of special significance in modulating thermal sensitivity. At the mechanistic level, where we analyze how specific substitutions exert their effects on thermal sensitivity of structure and function, we show that mobile regions, whose sequences are highly conserved, play a critical role in adaptation to temperature. Thus, substitutions that modify the energy barriers to movement of MRs appear to be of central importance in adaptation. These substitutions lie outside of the MRs and have their effects transmitted to the MRs through concerted interactions across the protein surface. In addition to intra-subunit effects, the sites that affect subunit-subunit interaction can also affect the stability of dimers. Thus, for multi-subunit proteins, it is necessary to take into account multimeric effects to understand fully the process of thermal adaptation.

Amino Acid Usage and Adaptation Temperature. Several significant trends between enzyme thermal stability and amino acid usage were revealed in this study (Fig. 2). These trends include a negative correlation between thermal stability and the percentage contributions of charged amino acids to the entire protein structure and to the protein surface (Fig. 2 *A, D*). The overall content of weakly hydrophobic and polar residues was significantly higher in warm-adapted orthologs, but no significant trends in individual protein regions were observed (Fig. 2 *B*–*K*). We also show that the frequency of occurrence of proline and glycine in the entire enzyme and on the protein surface correlated positively with adaptation temperature (Fig. 2 *M, N, Q, R*). The importance of surface-localized substitutions has been shown in other studies as well (4, 11–13); we discuss below how these surface changes can cause different thermal sensitivities in structure and function.

The trends in amino acid usage found in this study contrast with some of the patterns previously observed in bacterial and archaeal proteins. Thus, in contrast to our findings, proteins of thermophilic bacteria were shown to display a bias in amino acid composition in which charged residues are present in relatively large numbers and polar residues are

relatively scarce (14). The broadest analysis of microbial proteins involved comparisons of proteomes of bacteria and archaea with OGTs from 8 °C to ~100 °C (10). This study revealed the following significant trends: total protein-wide contributions of charged and hydrophobic residues increased and those of polar residues decreased with rising OGT. In surface regions and intermediate interior regions of structure (between surface and core) these same patterns were observed. However, the deep core of the proteins did not mirror these patterns.

Trends in the frequency of occurrence of glycine and proline merit close consideration in the context of providing a mechanistic explanation of temperature-adaptive change. Glycine content (Fig. 2 *Q*–*T*) did not follow the previously reported trend of increasing in cold-adapted orthologs (15, 16); rather, in the cMDHs, glycine content on the surface region was positively correlated with adaptation temperature. This trend is especially evident in orthologs of *Echinolittorina* congeners. The ortholog of the more heat-adapted species, *E. malaccana*, had glycines in two positions (48 and 114) occupied by serines in the cMDH of its congener, *E. radiata* (Figs. 1 and 3); these were the only differences in sequences between the two species. Hence, the replacement of two serines with glycines, which in principle would be expected to decrease stability due to the enhanced flexibility of structure associated with glycine residues (17), has led to an increase in stability instead. We conjecture that the surface-localized glycines in heat-adapted orthologs, e.g., that of *E. malaccana*, may facilitate organization of the folded structure into a more compact and temperature-resistant protein because of the small size of a glycyl residue. Thus, adaptive changes in localized regions of a protein may involve amino acid substitutions that disobey the canonical rules developed from analysis of entire proteomes.

The significant temperature-related trend in proline content observed for the 26 cMDH orthologs may account, at least in part, for the differences in resistance to heat denaturation and to structural perturbations that disrupt ligand binding. Thus, the relatively high rigidity in structure at and around proline residues may be important in reducing thermal sensitivity of protein conformation (17). Positive correlations between proline content and adaptation temperature have been reported in other studies (4, 18–21). A key role for proline in modulating thermal stability is suggested for cMDHs of two species of littorinid snails belonging to the genera *Echinolittorina* (*E. malaccana*) and *Littorina* (*L. keenae*); both species have relatively high upper thermal limits (57 °C and 48 °C, respectively. Table 1). The only two differences in amino acid sequences for cMDH orthologs of these two species occurred at sites 4 and 114: proline (site 4) and glycine (site 114) for *E. malaccana* and alanine residues at both positions for *L. keenae*. The importance of proline's stabilizing effects was shown in the studies employing SDM and MDS approaches, as discussed below.

Linking Sequence Variation to Thermal Stability of Structure and Function. To test conjectures about the roles of specific amino acid substitutions on structural stability and thermal sensitivity of cofactor binding, we conducted parallel *in silico* and *in vitro* studies on selected orthologs. These experiments focused on the two mobile regions, MR1 and MR2 (Fig. 4), which comprise residues 90–105 and 230–245, respectively (Fig. 1). We concentrated on MR1 and MR2 because of their pivotal roles in binding and catalysis and the

occurrence of significant inter-ortholog differences in their thermal sensitivities (2) despite a high degree of sequence conservation in both MRs (Fig. 1). As shown in Fig. 7, these two regions exhibit the highest rmsf values in the cMDH sequence. The magnitude of rmsf values in the MRs is commonly, although not invariably lower in orthologs of more warm-adapted congeners (compare values for *E. malaccana* and *E. radiata* (Fig. 7 *A, B*). Furthermore, the effects of increases in simulation temperature on rmsf values in the MRs are lower in warm-adapted than cold-adapted orthologs (2).

Identifying the specific amino acid substitutions that are responsible for the inter-ortholog differences in flexibility in MR1 and MR2 can be made with varying degrees of confidence. The clearest cause-effect linkage is for the two congeners of *Lottia*, whose cMDHs differ by only a single substitution: at position 291, the serine in the ortholog of the more heat-tolerant species, *L. austrodigitalis*, is replaced by a glycine in the ortholog of *L. digitalis* (Fig. 4) (22). Despite this minor variation in sequence, the number of residues in the MRs that show a change in rmsf in response to an increase in simulation temperature is smaller in *L. austrodigitalis* than in *L. digitalis* (Table 1 in ref. 2). Even though site 291 falls well outside of MR1 and MR2, it appears to affect the temperature sensitivity as indexed by the change in rmsf of both mobile regions, as well as movement of backbone structures throughout the protein, as indexed by rmsd (2). A similar pattern was observed between the cMDHs of the two congeners of *Echinolittorina*: the only two amino acid substitutions reside outside of the MRs yet they influence the flexibility of the MRs, and account for the variable thermal sensitivities of structure and cofactor binding between orthologs (3). We conjecture that the sites of this variation lie largely in regions of the sequence that influence the conformational stabilities of the MRs themselves and/or the energy changes that occur during formation of the ternary complex, notably the collapse of the catalytic loop during cofactor binding.

Previous studies have reached similar conclusions about the functionally important roles of substitutions at sites distant from active sites. For example, a study on α-amylase from an Antarctic bacterium pointed out the importance of effects on protein flexibility induced by amino acid substitutions at sites distant from the active site and substrate binding groove (12). Comparable effects of amino acid substitutions at positions distant from the active site have been found in studies of solute effects on proteins. Studies of adaptive evolution of a fish hatching enzyme, high choriolytic enzyme (HCE) showed that a single amino acid substitution results in differential salt dependency of orthologs of this enzyme. Superimposition of killifish HCEs with the 3D-structure model of medaka HCE revealed that the single substitution was located on the surface of HCE molecule, far from its active site cleft (23). In summary, all of these findings provide support for the importance of effects of transmitted conformational effects through large regions of the protein, in keeping with the analysis of Gerstein and Chothia (24) and other (25).

Whereas we have focused up to this point on structure-function effects seen at the level of the single subunit, certain of the substitutions we have found may influence the interactions between the two subunits of the cMDH dimer and alter the thermal sensitivity of the dimeric enzyme. Sequences of the subunit contact sites themselves do not show evidence for adaptive change (Fig. 1): the only sequence variation in the region of

subunit-subunit interaction is at site 22 (tyrosine, histidine or leucine). Our data again contrast with studies of microorganisms. Studies of thermophilic bacteria (5, 26) have proposed that the presence of increased numbers of charged residues at subunit contact sites fosters formation of additional ionic interactions across the dimer–dimer interface, which favors stability of the enzyme dimer (5). In an MDH from a thermophilic bacterium, the number of charged residues involved in stabilizing subunit-subunit interactions was higher than in cMDH from a mammal (5). These differences were conjectured to contribute to elevated thermal stability of the bacterial enzyme.

In our comparisons of warm- and cold-adapted cMDH orthologs, the conservative changes at site 22 would not lead directly to differences in ionic interactions for stabilizing the dimer. Thus, within the species of molluscs we examined, the observed differences in global thermal stability are likely not due to different strengths of inter-subunit interactions that arise from the amino acids present at the subunit-subunit interaction sites. However, the possibility remains that the strengths of these interactions could be influenced by differences in conformational stability of the subunits. These differences could arise from amino acid substitutions distant from subunit interaction sites yet which lead to stabilization of the dimer. These transmitted effects, wherein changes at sites distant from the site at which functionally important effects occur (here, stabilization of subunit-subunit interactions), may thus pertain for stabilizing the dimer as well as for stabilizing ligand binding. A conceptual framework for evaluating these types of transmitted effects has been developed for an enzyme closely related to MDH, lactate dehydrogenase (LDH) (24). In this model, the protein can be regarded as comprising "concentric shells of increasing mobility" (core to surface), with large conformational changes in surface regions being transmitted throughout much and perhaps all of the molecule (24). Temperature-induced conformational changes in relatively flexible surface regions of a protein thus could be transmitted to sites involved in subunit interactions, as well as to catalytic sites. Influences on the mobilities of MRs could be critical in both types of effects. This model can account for the findings that amino acid substitutions that affect thermal stability of the protein and its catalytic function (indexed by K_M^{NADH}) occur distant from the enzyme's active site and the subunit-subunit interface region.

How do the Findings of *in silico* and *in vitro* Analyses Compare? One of the purposes of this study was to determine whether mutagenesis involving SDM methods, followed by *in vitro* study of the over-expressed protein, and *in silico* methods using MDS led to similar conclusions about the role of specific amino acid residues in establishing protein thermal sensitivity. As shown by data in Figures 6–9, agreements between these two methods were found in some, but not all comparisons. For example, the mutation induced at site 4 in the cMDH of *L. keenae*—the replacement of an alanine by a proline, mut-A4P—led to a decrease in rmsd (Fig. 6 *E*), reflecting increase backbone rigidity, and a decrease in rate of denaturation *in vitro* (Fig. 9 *B*). In contrast to this consistency between the *in silico* and *in vitro* approaches, the replacement of the proline residue found at site 4 in the *E. malaccana* ortholog with an alanine did not lead to significant change in rmsd (Fig. 6 *D*) even though the *in vitro* thermal stability of the enzyme (mut-P4A) significantly decreased relative to recombinant wild-type enzyme (wt-*Em*) (Fig. 9 *B*). The results of the rmsf analyses (Fig. 7)

may provide some insights into these disparate results. In the case of the wt-*Lk* versus mut-A4P comparison (Fig. 7 *D*) the mobility of residues near site 50 in the sequence was greatly reduced in the more stable mut-A4P variant of the enzyme. This region lies near the other subunit's MR2 in the dimer, so the alanine to proline change might lead to improved monomer-monomer contacts in the dimer as well as to effects on the mobility of MR2. Thus, it seems possible that the conformational flexibility of the region near site 50 could be critical in governing the thermal responses of the native dimeric protein. Shifts in the rmsf values of residues in MR1 are also seen, especially in residues near sites 95–100. In the comparison of wt-*Em* and mut-P4A, a large number of residues throughout the sequence exhibited shifts in rmsf value (Fig. 7 *C*). However, how these changes in flexibility translate into a decrease in thermal stability cannot be determined with confidence.

In the cMDHs of the two *Echinolittorina* congeners, consistencies and inconsistencies were again found between *in silico* and SDM-*in vitro* studies that involved mutations at site 114. There was again a type of asymmetry in the observed changes. Thus, for the ortholog of *E. malaccana,* replacing the glycine residue at site 114 with a serine (wt-*Em* to mut-G114S) did not cause a change in rmsd (Fig. 6 *A*) or alter thermal sensitivity *in* vitro (Fig. 9 *A*). However, doing the opposite mutation in the ortholog of *E. radiata*—replacing the serine with a glycine (wt-*Er* to mut-S114G)—significantly increased rmsd (Fig. 6 *B*) and the rate of loss of activity during heating (Fig. 9 *A*). A correlate of these effects on rmsd and *in vitro* stability is seen in the rmsf data (Fig. 7). The thermally sensitive mut-S114G enzyme exhibits a large increase in residue mobility near sites 45-50, the region of sequence that showed a large decrease in mobility in the mut-A4P enzyme, which had enhanced *in vitro* stability (Fig. 9 *B*).

A further comparison of relevance contrasts the effects of mutations on *in vitro* thermal stability (Fig. 9) and ligand binding (K_M^{NADH}) (Fig. 8). In some cases, a change in resistance to thermal denaturation was accompanied by altered thermal perturbation of binding. For example, mut-P4A showed greater thermal sensitivity in binding as well as reduced protein stability relative to wt-*Em*. A similar agreement was seen for the comparison of wt-*Er* and mut-S114G. However, whereas wt-*Em* and mut-G114S showed no difference in thermal stability of structure, a significantly increased sensitivity of K_M^{NADH} was found in the mut-G114S variant. The uncoupling of effects on structure and binding seen in the latter case indicates that a mutation that alters temperature-sensitivity of enzyme-ligand interactions need not alter the global stability of the protein.

In summary, it is noteworthy that the results of the MDS experiments mirror the trends of SDM-*in vitro* studies only in part. As emphasized in a recent review, simulation methods may fall short of *in vitro* methods in certain instances due largely to current limitations in *in silico* algorithms (27). In the present study, the differences between the *in silico* and *in vitro* analyses in thermal sensitivity of the different variants of cMDH may be a consequence of a focus on the monomer in the *in silico* studies and the use of a dimeric enzyme in the *in vitro* work. Thus, amino acid substitutions that affect the stability of the dimer would not necessarily lead to changes in, for example, the computed rmsf values of a simulated monomer. Further parallel use of these *in silico* and SDM-*in vitro* approaches are needed, to

allow a more comprehensive analysis of the potential utility of MDS methods for testing putative adaptive significance of amino acid substitutions. At the very least, however, MDS methods are helpful in identifying sites in the protein sequence where changes in conformational flexibility are induced by an amino acid substitution. The present studies illustrate this contribution of MDS approaches and show how the identification of the sites of changes in conformational flexibility leads to a deeper understanding of how regions of surface structure distant to protein active sites can contribute to the functional properties of the protein. Combined studies using both *in silico* MDS methods and *in vitro* site-directed mutagenesis thus can provide a clearer understanding of thermal adaptation of proteins and protein structure-function relationships in general.

Materials and Methods

Sequencing of cMDH cDNA. The cMDH orthologs that had not been sequenced previously were sequenced with the protocols described by Dong and Somero (22). The full-length cDNA sequence was obtained using the 5'- and 3'-rapid amplification of cDNA ends (RACE) protocol (Clontech Laboratories, Inc., CA, USA) and sent to a commercial company for sequencing (Sangon Biotech Co., Ltd., Shanghai, China). The primers designed for sequencing and the GenBank accession numbers of cMDH cDNAs are shown in SI Appendix, Table S5 and Table S6, respectively.

Analysis of Overall- and Region-Specific Variation in Amino Acid Composition. Using the sequence data of cMDH cDNAs, three-dimensional models were constructed by the I-TASSER server that had high C-scores (1.5–1.6) (28). C-score is a confidence score for estimating the quality of predicted models, which is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations (28). C-score is typically in the range between −5 and 2, where a higher (positive) value signifies higher confidence of the model. The VMD program was used to calculate the solvent accessible surface area based on the computed three-dimensional structure (29). Relative accessible surface area (RSA) was calculated as described by Tien et al. (30) to classify positions of residues. Residues with zero RSA were categorized as core residues, residues with RSA >25% were categorized as surface residues, and the remaining residues were assigned to the intermediate region.

Molecular Dynamic Simulation of cMDH. The computed three-dimensional structures above were used as the starting models of the simulations. Simulations were performed using NAMD 2.9 (31) as described by Dong et al. (2) for 20 ns at 57 °C (N=5 or 10). Because the values of the rmsd of backbone atom positions stabilized after 10 ns, 20 ns was selected as a suitable and economical timeframe for the MDS of molluscan cMDH based on the test simulations (SI Appendix, Fig. S4). The VMD program was used to visualize and analyze the simulation trajectories (29). The rmsd of backbone atom positions and the rmsf for individual residues in all models were calculated and compared. For rmsd and rmsf calculations, the initial and energy-minimized structures, respectively, were used as the reference. They are defined as: $\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i - r_0)^2}$ and

$\mathrm{RMSF} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i - r)^2}$, where r_i represents the position at time *i*, r_0 represents the reference value, r represents the average value of the root mean square fluctuation and N represents the number of atoms. The stabilized structure (10–20 ns) was taken from the trajectory of the system to determine the movements of protein backbone and individual residue atoms. Differences among cMDH orthologs simulations were analyzed using one-way analysis of variance (ANOVA) followed by the Tukey's multiple comparisons test (*P* = 0.05) with Prism 6.0 software (GraphPad Software, La Jolla, USA).

Site-directed Mutagenesis, Expression, and Purification of Recombinant Enzymes. cMDHs cDNAs were amplified by polymerase chain reaction (PCR) using pairs of primers with restriction sites for EcoRI and XhoI (SI Appendix, Table S5). PCR products were purified and digested by EcoRI and XhoI, and then ligated to pMD19 (TaKaRa, Dalian, China) to generate the recombinant plasmids using T4 DNA ligase. The *E. coli* DH5α (TaKaRa, Dalian, China) was used for initial cloning of pMD19-cMDH. Pairs of primers were designed for the SDM with the TaKaRa MutanBEST Kit (Dalian, China). The recombinant plasmids pMD19-mut-P4A and pMD19-wt-*Em* were used as the templates to generate the recombinant plasmids pMD19-wt-*Lk* and pMD19-mut-A4P, respectively. All the pMD19-cMDHs were purified and digested by EcoRI and XhoI, and then cloned into vectors pET32a(+) that digested by the same restriction enzyme above. The pET32a(+) recombinant vectors containing either the wild types (wt-*Em*, wt-*Er*, and wt-*Lk*), or mutants types (mut-G114S, mut-S114G, mut-P4A, and mut-A4P) were transformed into expression hosts, *E. coli* BL21(DE3) chemically competent cells (Weidi, Shanghai, China) and plated on Luria-Bertani (LB) agar medium containing ampicillin (100 μg/mL) at 37 °C overnight.

A single colony of transformed *E. coli* BL21(DE3) with pET32a(+)-cMDH was incubated overnight on a shaking incubator (Yiheng Scientific In-strument Co., Ltd.; THZ-103B, with a rotational radius of 20 cm) in 10 mL LB/ampicillin broth medium at 37 °C with constant agitation (180 rpm). Plasmids were purified from individual colonies using an extraction kit (Aidlab, Beijing, China). The sequences of target cMDHs in plasmids were verified by DNA sequencing. 200 mL of LB/ampicillin broth was inoculated and grown to log phase with vigorous shaking (180 rpm) at 37 °C for ~6h. Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.5 mmol/L for expressions of target fusion proteins. The incubation period continued for another six hours at 30 °C with shaking at 260 rpm. After inductions, the bacteria cells were pelleted at 8000 g for 10 min. The expressed protein was purified using an Ni-NTA agarose column according to manufacturer's instruction (Smart-lifesciences, Changzhou, China) at 4 °C. The pellet was re-suspended in 10 mL lysis buffer (50 mmol/L potassium phosphate, pH 7.0, 300 mmol/L NaCl). Extraction was performed by an ultrasonic cell disruptor (Branson, Danbury, USA) with 30% amplitude kept cold by an ice-water mixture (an extraction cycle with 10 s disruption and 10 s rest, totally 30 cycles of 10 min). The cell suspension was then centrifuged at 12,000 g and 4 °C for 60 min. The supernatant was filtered (0.45 μm) and then passed over a 5-mL His-tag affinity column, which bound the poly His-tag attached to the N-terminus of the recombinant cMDH and allowed rapid purification. After the column was washed (50 mL of 50 mmol/L potassium phosphate, pH 7.0, 300 mmol/L NaCl), the purified cMDH was eluted with 10 mL wash buffer plus 150 mmol/L imidazole. The purified protein was dialyzed with potassium phosphate

buffer (50 mmol/L, pH 6.8) at 4 °C overnight. The qualities of purified recombinant proteins were analyzed on SDS-PAGE gel electrophoresis.

Determination of K_M^{NADH} and Rate of Heat Denaturation. The purified and dialyzed proteins above were used for assays. The reaction mixture for determining K_M^{NADH} contained imidazole-Cl buffer (200 mmol/L, pH 7.0 @ 25 °C), oxaloacetic acid (200 µmol/L), and different concentrations of NADH (10, 20, 40, 75, 100, and 150 µmol/L). K_M^{NADH} values were calculated from the initial velocities of the reaction at each NADH concentration using Prism 6.0 software (GraphPad Software, La Jolla, USA). Thermal stabilities were carried out as described by Fields et al. (32). Dialyzed enzyme preparations were heated at 40° or 57.5 °C for different periods of time. Aliquots were removed and assayed for residual cMDH activity. Differences among values were analyzed using one-way analysis of variance (ANOVA) followed by the Tukey's multiple comparisons test (P = 0.05) with Prism 6.0 software (GraphPad Software, La Jolla, USA).

ACKNOWLEDGMENTS. We thank Mr. Yu-wei Yu and Mr. Zhuo-bin Xu from the Information and Network Center, Xiamen University, for their help with high-performance computing. We thank Dr. Donal Manahan for helpful collection of the Antarctic specimens. We thank the support from the State Key Laboratory of Marine Environmental Science, Xiamen University. This work was supported by National Natural Science Foundation of China Grants 41776135 and 41476115 (to Y.-w.D.) and Nature Science Funds for Distin-guished Young Scholars of Fujian Province, China Grant 2017J07003 (to Y.-w.D.). Work in the collection of the Antarctic specimens was supported by National Science Foundation Grant 1245703 (to Donal Manahan, co-PI).

References

1. Somero GN, Lockwood BL, Tomanek L (2017) *Biochemical Adaptation: Response to Environmental Challenges from Life's Origins to the Anthropocene* (Sinauer Associates, Sunderland, MA).
2. Dong YW, Liao ML, Meng XL, Somero GN (2018) Structural flexibility and protein adaptation to temperature: Molecular dynamics analysis of malate dehydrogenases of marine molluscs. *Proc Natl Acad Sci USA* 115:1274–1279.
3. Liao ML, et al. (2017) 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* 220:2066–2075.
4. Fields PA, Dong YW, Meng XL, Somero GN (2015) Adaptations of protein structure and function to temperature: there is more than one way to 'skin a cat'. *J Exp Biol* 218:1801–1811.
5. Dalhus B, et al. (2002) Structural basis for thermophilic protein stability: structures of thermophilic and mesophilic malate dehydrogenases. *J Mol Biol* 318:707–721.
6. Chapman AD, et al. (1999) Structural basis of substrate specificity in malate dehydrogenases: crystal structure of a ternary complex of porcine cytoplasmic malate dehydrogenase, α-ketomalonate and tetrahydoNAD. *J Mol Biol* 285:703–712.
7. Birktoft JJ, Rhodes G, Banaszak LJ (1989) Refined crystal structure of cytoplasmic malate dehydrogenase at 2.5-Å resolution. *Biochemistry* 28:6065–6081.

8. Karshikoff A, Nilsson L, Ladenstein R (2015) Rigidity versus flexibility: the dilemma of understanding protein thermal stability. *Febs Journal* 282:3899–3917.
9. Siddiqui KS (2017) Defying the activity-stability trade-off in enzymes: taking advantage of entropy to enhance activity and thermostability. *Crit Rev Biotechnol* 37:309–322.
10. Ma BG, Goncearenco A, Berezovsky IN (2010) Thermophilic adaptation of protein complexes inferred from proteomic homology modeling. *Structure* 18:819–828.
11. Isaksen GV, Aqvist J, Brandsdal BO (2014) Protein surface softness is the origin of enzyme cold-adaptation of trypsin. *PloS Comput Biol* 10:e1003813.
12. Papaleo E, Pasi M, Tiberti M, De Gioia L (2011) Molecular dynamics of mesophilic-like mutants of a cold-adapted enzyme: insights into distal effects induced by the mutations. *PLoS One* 6:e24214.
13. Pasi M, et al. (2009) Dynamic properties of a psychrophilic alpha-amylase in comparison with a mesophilic homologue. *J Phys Chem B* 113:13585–13595.
14. Fukuchi S, Nishikawa K (2001) Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria. *J Mol Biol* 309:835–843.
15. Gerday C, et al. (2000) Cold-adapted enzymes: from fundamentals to biotechnology. *Trends Biotechnol* 18:103–107.
16. Haney PJ, et al. (1999) Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. *Proc Natl Acad Sci USA* 96:3578–3583.
17. Matthews BW (1987) Genetic and structural analysis of the protein stability problem. *Biochemistry* 26:6885–6888.
18. Yang LL, Tang SK, Huang Y, Zhi XY (2015) Low temperature adaptation is not the opposite process of high temperature adaptation in terms of changes in amino acid composition. *Genome Biol Evol* 7:3426–3433.
19. Goihberg E, et al. (2007) A single proline substitution is critical for the thermostabilization of *Clostridium beijerinckii* alcohol dehydrogenase. *Proteins: Struct Funct Bioinform* 66:196–204.
20. Fields PA, Somero GN (1998) Hot spots in cold adaptation: localized increases in conformational flexibility in lactate dehydrogenase A_4 orthologs of Antarctic notothenioid fishes. *Proc Natl Acad Sci USA* 95:11476–11481.
21. Wang GZ, Lercher MJ (2010) Amino acid composition in endothermic vertebrates is biased in the same direction as in thermophilic prokaryotes. *BMC Evol Biol* 10:263.
22. Dong YW, Somero GN (2009) 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* 212:169–177.
23. Kawaguchi M, et al. (2013) Adaptive evolution of fish hatching enzyme: one amino acid substitution results in differential salt dependency of the enzyme. *J Exp Biol* 216:1609–1615.
24. Gerstein M, Chothia C (1991) Analysis of protein loop closure. Two types of hinges produce one motion in lactate dehydrogenase. *J Mol Biol* 220:133–149.
25. Holland LZ, McFall-Ngai M, Somero GN (1997) 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* 36:3207–3215.
26. Kelly CA, et al. (1993) Determinants of protein thermostability observed in the 1.9-Å crystal structure of malate dehydrogenase from the thermophilic bacterium *Thermus flavus*. *Biochemistry* 32:3913–3922.

27. Bottaro S, Lindorff-Larsen K (2018) Biophysical experiments and biomolecular simulations: A perfect match. *Science* 361:355–360.
28. Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9:40.
29. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:33–38.
30. Tien MZ, et al. (2013) Maximum allowed solvent accessibilites of residues in proteins. *PLoS One* 8:e80635.
31. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781–1802.
32. Fields PA, Rudomin EL, Somero GN (2006) 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* 209:656–667.
33. Li HT (2012) Thermal tolerance of *Echinolittorina* species in Hong Kong : implications for their vertical distributions. Master of Philosophy (The University of Hong Kong, Hong Kong).
34. Somero GN (2002) Thermal physiology and vertical zonation of intertidal animals: optima, limits, and costs of living. *Integr Comp Biol* 34:67–75.
35. Miller LP, Harley CDG, Denny MW (2009) The role of temperature and desiccation stress in limiting the local-scale distribution of the owl limpet, *Lottia gigantea*. *Funct Ecol* 23:756–767.
36. Dong Y, Miller LP, Sanders JG, Somero GN (2008) Heat-shock protein 70 (Hsp70) expression in four limpets of the genus *Lottia*: interspecific variation in constitutive and inducible synthesis correlates with in situ exposure to heat stress. *Biol Bull* 215:173–181.
37. Tomanek L, Somero GN (1999) Evolutionary and acclimation-induced variation in the heat-shock responses of congeneric marine snails (genus *Tegula*) from different thermal habitats: implications for limits of thermotolerance and biogeography. *J Exp Biol* 202:2925–2936.
38. Chen N, et al. (2016) Assessment of the thermal tolerance of abalone based on cardiac performance in *Haliotis discus hannai*, *H. gigantea* and their interspecific hybrid. *Aquaculture* 465:258–264.
39. Cheng W, Hsiao IS, Hsu CH, Chen JC (2004) Change in water temperature on the immune response of Taiwan abalone *Haliotis diversicolor supertexta* and its susceptibility to *Vibrio parahaemolyticus*. *Fish Shellfish Immunol* 17:235–243.
40. Dahlhoff EA, Somero GN (1993) Kinetic and structural adaptations of cytoplasmic malate dehydrogenases of eastern Pacific abalone (genus *Haliotis*) from different thermal habitats: biochemical correlates of biogeographical patterning. *J Exp Biol* 185:137–150.
41. Jenewein BT, Gosselin LA (2013) Ontogenetic shift in stress tolerance thresholds of *Mytilus trossulus*: effects of desiccation and heat on juvenile mortality. *Mar Ecol Prog Ser* 481:147–159.
42. Denny MW, Dowd WW, Bilir L, Mach KJ (2011) Spreading the risk: Small-scale body temperature variation among intertidal organisms and its implications for species persistence. *J Exp Mar Biol Ecol* 400:175–190.
43. Peck LS, Webb KE, Bailey DM (2004) Extreme sensitivity of biological function to temperature in Antarctic marine species. *Funct Ecol* 18:625–630.
44. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157:105–132.

Figure legends

Fig. 1. Deduced amino acid sequence alignments of 26 cMDH orthologs from marine molluscs. The substrate-binding sites, the cofactor binding sites, the active site proton acceptor site, and the residues involved in subunit-subunit interaction are shown in red circles, blue triangles, a green pentagon, and orange diamonds, respectively. Conserved amino acid residues are highlighted in bold. Non-conservative replacements among *Echinolittorina malaccana*, *E. radiata*, *Littorina keenae*, and *L. scutulata* cMDHs are highlighted in red. Mobile regions (MRs) are highlighted in red frames.

Fig. 2. Overall and region-specific variation in different categories of amino acids of 26 orthologs of molluscan cMDHs as a function of enzyme stability (rate of denaturation at 40 °C for genus *Haliotis*, *Adamussium colbecki*, and *Laternula elliptica*, and at 42.5 °C for all other species). The fitted linear regression line is shown as a solid line when $P < 0.05$; a dashed line is used when the regression was non-significant. Amino acid categories were: charged (DEKR), weakly hydrophobic/polar (AGNQSTHY), and hydrophobic (LVWIFMPC) (44). (*A*–*C*) Overall variation in amino acid composition throughout the protein structure, (*D*–*F*) variation in surface composition, (*G*–*I*) in composition of the intermediate depth region, and (*J*–*L*) in composition of the core region of the protein. Variation in proline (*M*–*P*) and glycine (*Q*–*T*) content in different regions of the enzyme.

Fig. 3. Relationship between median lethal temperature (LT_{50}) and the rate of denaturation (slope of ln residual activity) of cMDH orthologs from marine molluscs. Each point represents the LT_{50} from one species. Regression coefficients are: $Y = 0.00603*X - 0.279$, $R^2 = 0.720$ (linear regression) and $Y = -0.391 + 0.0155*X – 0.000154*X^2$, $R^2 = 0.866$ (quadratic regression).

Fig. 4. Three-dimensional models of a single cMDH monomer from *Echinolittorina malaccana*, *E. radiata*, *Lottia austrodigitalis*, *L. digitalis*, *Chlorostoma funebralis*, and *C. montereyi*. Red-colored ribbons identify the regions in which the increase in simulation temperature led to a significant change in structural movements [indexed by a change in root mean square fluctuation (ΔRMSF) over the equilibration state (10–20 ns) greater than 0.5 Å. Simulation temperatures were 15 °C and 57 °C for the genus *Echinolittorina*, and 15 °C and 42 °C for the other orthologs (see reference 2). The locations of the two mobile regions are shown by dashed lines in *E. malaccana* cMDH monomer. The variable sites between species within each genus are indicated by yellow spheres.

Fig. 5. Dimeric assembly of *Echinolittorina malaccana* cMDH. (*A*) The substrate-binding sites, the cofactor binding sites, the active site proton acceptor, and the residues involved in subunit-subunit interaction are shown in red, blue, mauve, and yellow spheres, respectively.

The locations of the two mobile regions (residues 90–105 for MR1 and residues 230–245 for MR2) are highlighted by red shading. (*B*) Non-conservative replacements among *E. malaccana*, *E. radiata*, *Littorina keenae*, and *L. scutulata* cMDHs are shown in cyan spheres (residues 4, 33, 41, 48, 114, 219, and 321). The locations of the two MRs are shown by red dashed lines.

Fig. 6. The rmsd of backbone atom positions for cMDHs at a simulation temperature of 57 ° C. (*A* and *B*): Two wild type cMDHs, *Echinolittorina malaccana* (wt-*Em*) and *E. radiata* (wt-*Er*), and two mutant forms, mut-G114S (glycine mutated to serine at site 114 for the *E. malaccana* ortholog) and mut-S114G (serine mutated to glycine for the *E. radiata* ortholog). (*C*): the corresponding equilibrium rmsd values (10–20 ns as shaded) (N=10) and statistical analyses. (*D* and *E*): two wild type enzymes, wt-*Em* and *Littorina keenae* (wt-*Lk*), and two mutant enzymes, mut-P4A (proline mutated to alanine at site 4 for *E. malaccana* ortholog) and mut-A4P (alanine mutated to proline for *L. keenae* ortholog). (*F*): the corresponding equilibrium rmsd values (N=5) and statistical analysis. Data are expressed as means ± SEM and significance is denoted with asterisks: * for *P* value less than 0.05, ** for *P* values less than 0.01, and *** for *P* value less than 0.001.

Fig. 7. The root mean square fluctuation (rmsf) for individual residues over the equilibration state (10–20 ns) of cMDHs at a simulation temperature of 57 °C. (*A* and *B*) Two wild type enzymes, *Echinolittorina malaccana* (wt-*Em*) and *E. radiata* (wt-*Er*), and two mutant enzymes, mut-G114S (glycine mutated to serine at site 114 for the *E. malaccana* ortholog) and mut-S114G (serine mutated to glycine for the *E. radiata* ortholog) (N=10); (*C* and *D*) two wild type enzymes, wt-*Em* and *Littorina keenae* (wt-*Lk*), and two mutant enzymes, mut-P4A (proline mutated to alanine at site 4 for the *E. malaccana* ortholog) and mut-A4P (alanine mutated to proline for the *L. keenae* ortholog) (N=5). The locations of the two mobile regions (residues 90–105 for MR1 and residues 230–245 for MR2) are highlighted by gray shading.

Fig. 8. Temperature-dependence of the Michaelis-Menten constant for cofactor (reduced nicotinamide adenine dinucleotide) (K_M^{NADH}) for recombinant wild type and mutated cMDHs. (*A*) Two recombinant wild types, *Echinolittorina malaccana* (wt-*Em*) and *E. radiata* (wt-*Er*), and two mutants, mut-G114S (glycine mutated to serine at site 114 for the *E. malaccana* ortholog) and mut-S114G (serine mutated to glycine for the *E. radiata* ortholog); (*B*) two recombinant wild types, wt-*Em* and *Littorina keenae* (wt-*Lk*), and two mutants, mut-P4A (proline mutated to alanine at site 4 for *E. malaccana* ortholog) and mut-A4P (alanine mutated to proline for *L. keenae* ortholog). Inset: Arrhenius plot (ln K_M^{NADH} versus reciprocal temperatures in Kelvins) of cMDHs. Data are expressed as means ± SEM (N=5).

Fig. 9. Thermal stabilities of recombinant and wild type cMDH determined as the residual activities following incubation at 57.5 °C for different times. (*A*) two recombinant wild types, *Echinolittorina malaccana* (wt-*Em*) and *E. radiata* (wt-*Er*), and two mutants, mut-G114S (glycine mutated to serine at site 114 for the *E. malaccana* ortholog) and mut-S114G (serine mutated to glycine for the *E. radiata* ortholog); (*B*) two recombinant wild types, wt-*Em* and *Littorina keenae* (wt-*Lk*), and two mutants, mut-P4A (proline mutated to alanine at site 4 for the *E. malaccana* ortholog) and mut-A4P (alanine mutated to proline for the *L. keenae* ortholog). Inset: Slope of the linear regression of ln residual activity. Significance denoted with asterisks, * for *P* value less than 0.05, ** for *P* value less than 0.01, and *** for *P* value less than 0.001. Data are expressed as means ± SEM (N=3).

Table 1. Taxonomy and thermal tolerances [indicated as the median lethal temperature (LT_{50})] of the study species and the rate of denaturation (slope of ln residual activity) of the 26 molluscan orthologs of cMDH.

Taxonomy			Thermal tolerance		Reference
Family	Genus	Species	LT_{50} (°C)	Rate of denaturation of cMDH	
Littorinidae	*Echinolittorina*	*E. radiata*	55.5	-0.0007	3, 33
		E. malaccana	56.5	-0.0009	
	Littorina	*L. keenae*	47.5	-0.0033	2, 34
		L. scutulata	45.5	-0.0043	
Lottiidae	*Nipponacmea*	*N. fuscoviridis*	44.2	-0.0028	Present study
		N. radula	42.4	-0.0054	
	Lottia	*L. gigantea*	37.4	-0.0059	22, 35, 36
		L. austrodigitalis	41.1	-0.0098	
		L. scabra	-	-0.0124	
		L. digitalis	40.1	-0.0161	
		L. scutum	-	-0.0203	
		L. pelta	-	-0.0229	
Tegulidae	*Chlorostoma*	*C. rugosa*	45.0	-0.0089	2, 37
		C. funebralis	42.5	-0.0104	
		C. brunnea	36.0	-0.0166	
		C. montereyi	36.0	-0.0239	
Neritidae	*Nerita*	*N. yoldii*	50.0	-0.0004	2
		N. albicilla	51.0	-0.0010	
Haliotidae	*Haliotis*	*H. rufescens*	-	-0.0220	38–40
		H. diversicolor	31.1	-0.0667	
		H. discus	32.3	-0.1522	
Mytilidae	*Mytilus*	*M. trossulus*	34.1	-0.0395	32, 41, 42, present study
		M. californianus	38.2	-0.0417	
		M. galloprovincialis	37.3	-0.0602	
Laternulidae	*Laternula*	*L. elliptica*	8.3	-0.2296	43, present study
Pectinidae	*Adamussium*	*A. colbecki*	4.0	-0.3800	

Fig. 1

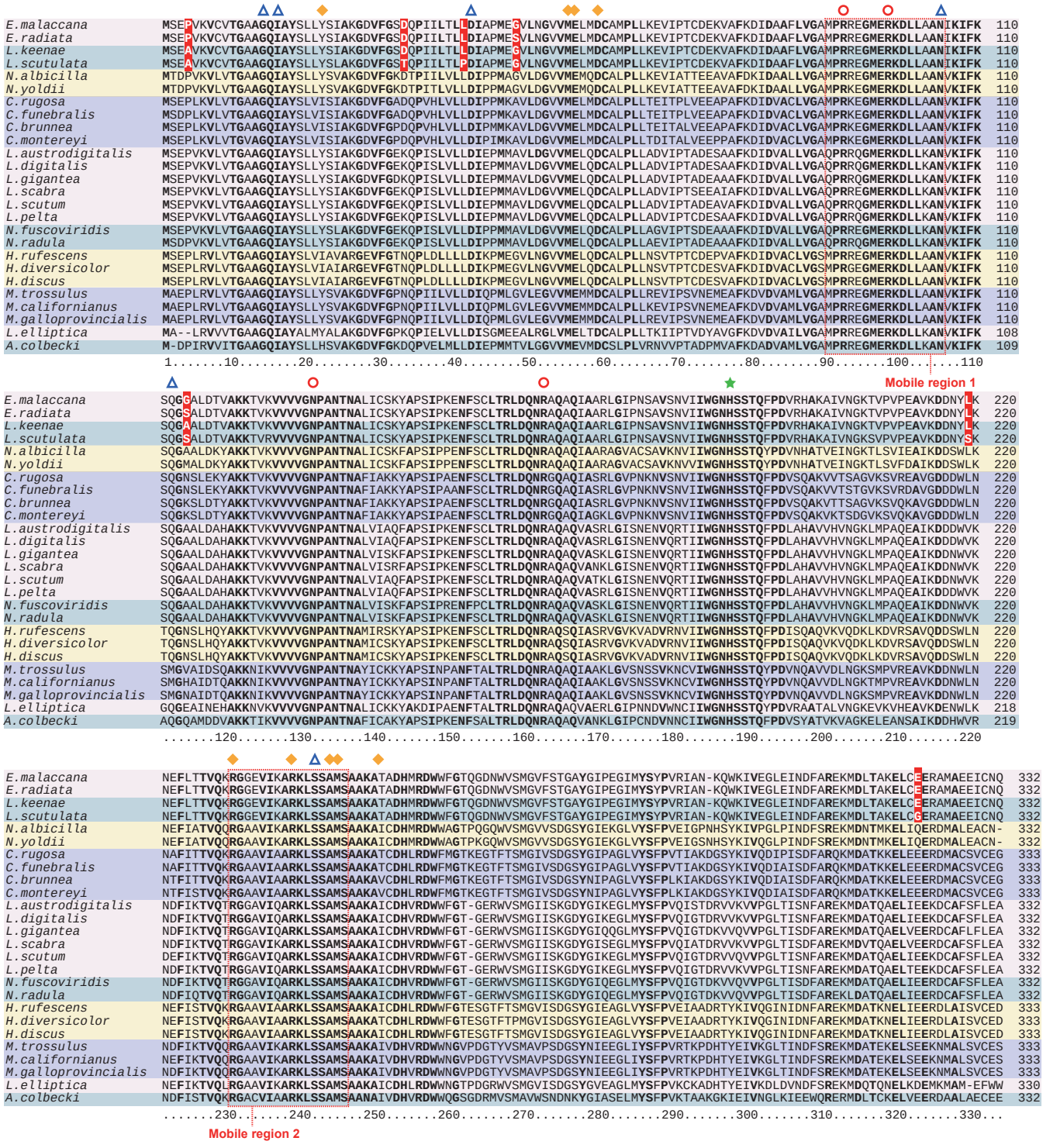


Fig. 2

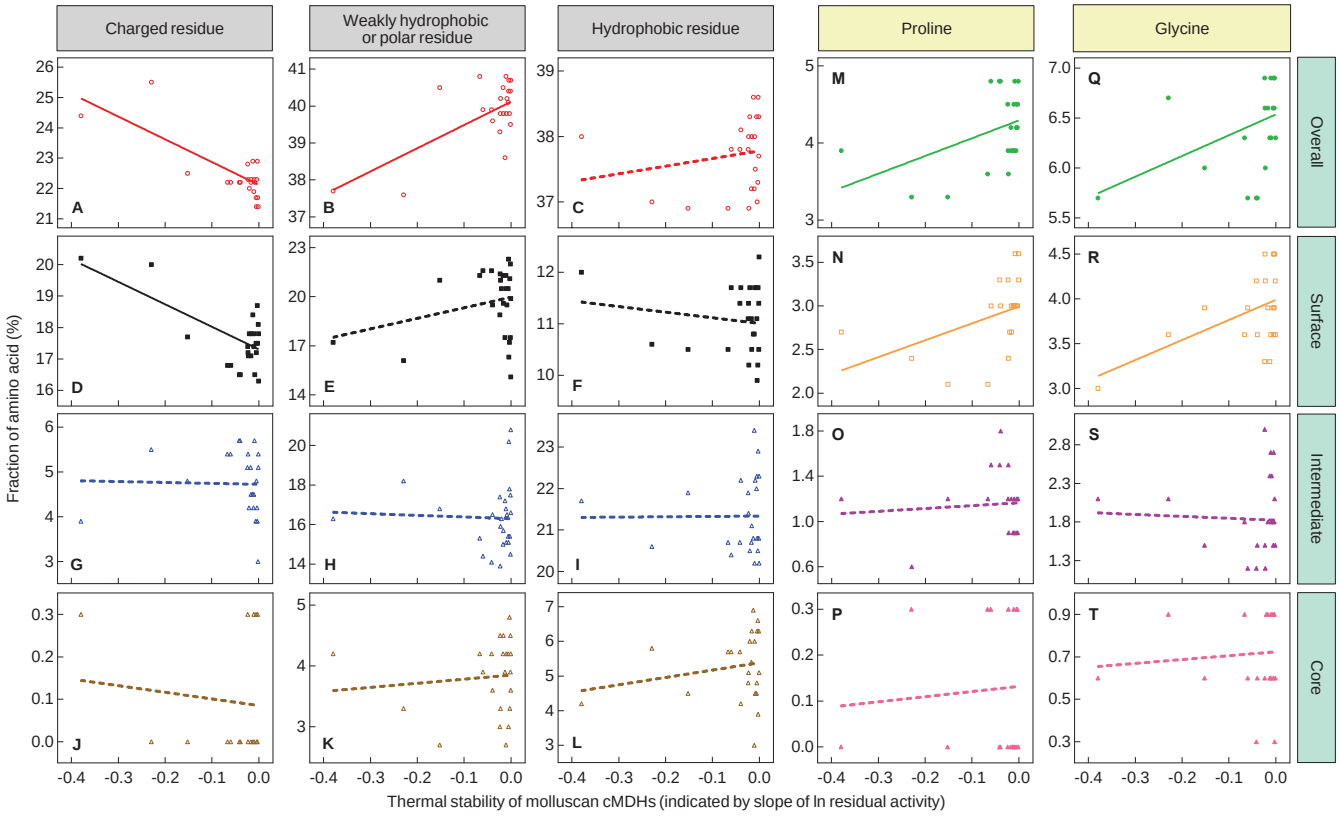

Charged residue
Weakly hydrophobic or polar residue
Hydrophobic residue
Proline
Glycine
Overall
Surface
Intermediate
Core
Fraction of amino acid (%)
Thermal stability of molluscan cMDHs (indicated by slope of ln residual activity)
A
B
C
M
Q
D
E
F
N
R
G
H
I
O
S
J
K
L
P
T

Fig. 3

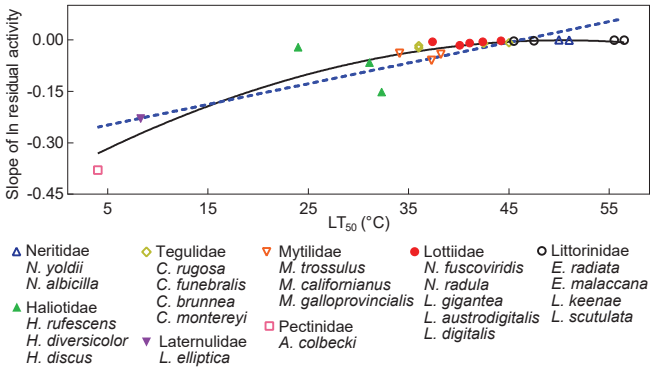

Slope of ln residual activity
0.00
-0.15
-0.30
-0.45
5
15
25
35
45
55
LT50 (°C)
Neritidae
N. yoldii
N. albicilla
Haliotidae
H. rufescens
H. diversicolor
H. discus
Tegulidae
C. rugosa
C. funebralis
C. brunnea
C. montereyi
Laternulidae
L. elliptica
Mytilidae
M. trossulus
M. californianus
M. galloprovincialis
Pectinidae
A. colbecki
Lottiidae
N. fuscoviridis
N. radula
L. gigantea
L. austrodigitalis
L. digitalis
Littorinidae
E. radiata
E. malaccana
L. keenae
L. scutulata

Fig. 4

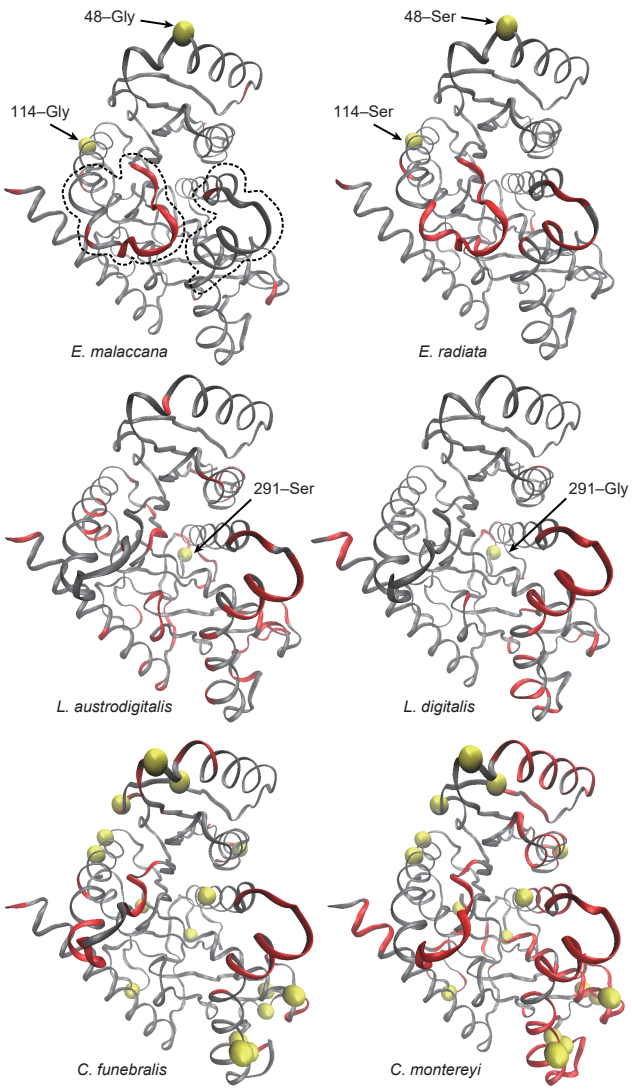

48–Gly
114–Gly
E. malaccana
48–Ser
114–Ser
E. radiata
291–Ser
L. austrodigitalis
291–Gly
L. digitalis
C. funebralis
C. montereyi

Fig. 5

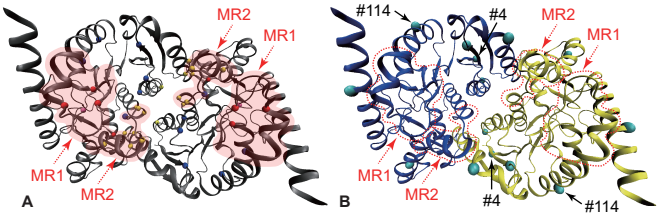


Fig. 6

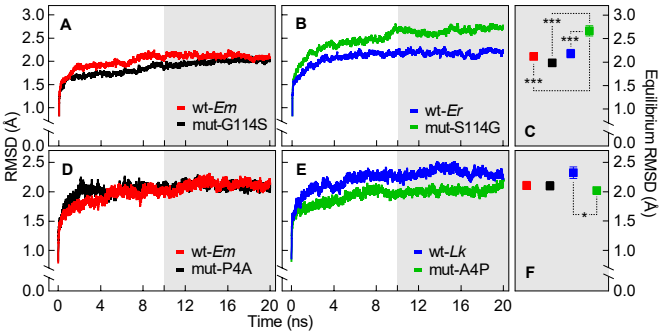

A
wt-Em
mut-G114S
B
wt-Er
mut-S114G
C
D
wt-Em
mut-P4A
E
wt-Lk
mut-A4P
F
RMSD (Å)
Equilibrium RMSD (Å)
Time (ns)
0.0
1.0
1.5
2.0
2.5
3.0
0
4
8
12
16
20
***
*

Fig. 7

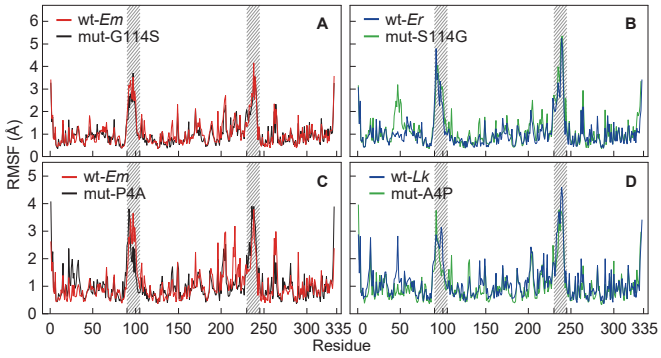

wt-Em
mut-G114S
A
wt-Er
mut-S114G
B
wt-Em
mut-P4A
C
wt-Lk
mut-A4P
D
RMSF (Å)
Residue
0
50
100
150
200
250
300
335

Fig. 8

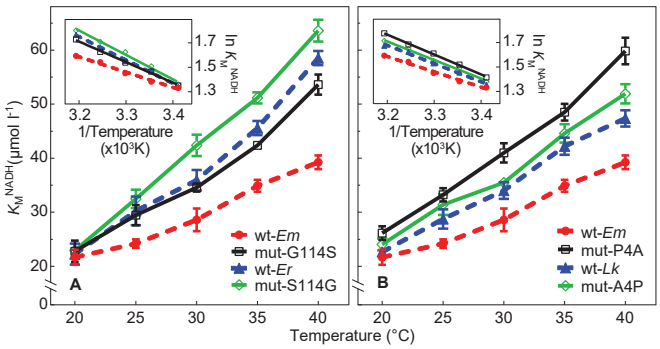
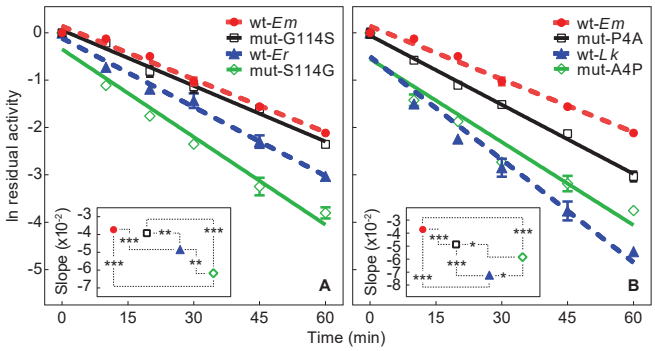

K_M^{NADH}(μmol l^{-1})
Temperature (°C)
1/Temperature (x10^{3}K)
ln K_M^{NADH}
wt-*Em*
mut-G114S
wt-*Er*
mut-S114G
mut-P4A
wt-*Lk*
mut-A4P
A
B

Fig. 9


wt-Em
mut-G114S
wt-Er
mut-S114G
wt-Em
mut-P4A
wt-Lk
mut-A4P
ln residual activity
Time (min)
Slope (x10-2)
A
B