PNAS
www.pnas.org

Supplementary Information for

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

Ming-ling Liao, George N. Somero, and Yun-wei Dong

Yun-wei Dong
Email: dongyw@xmu.edu.cn

This PDF file includes:

Figs. S1 to S4
Tables S1 to S6

www.pnas.org/cgi/doi/10.1073/pnas.1817455115

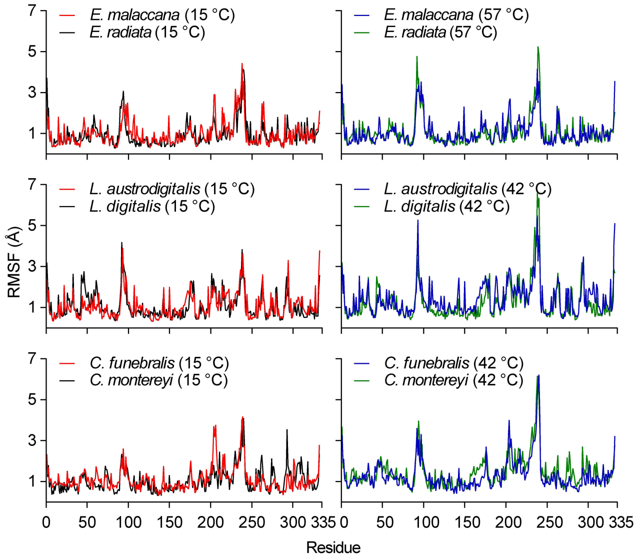

Fig. S1. The rms fluctuation (rmsf) for individual residues over the equilibration state (10–20 ns) of cMDHs from *Echinolittorina malaccana*, *E. radiata*, *Lottia austrodigitalis*, *L. digitalis*, *Chlorostoma funebralis*, and *C. montereyi*. Simulation temperatures: For genus *Echinolittorina*, 15 ℃ and 57 ℃, and for the other orthologs, 15 ℃ and 42 ℃.

```
wt-Em        MSEPVKVCVTGAAGQIAYSLLYSIAKGDVFGSDQPIILTLLDIAPMEGVLNGVVMELMDCAMPLLKEVIPTCDEKVAFKDIDAAFLVGAMPRREGMERKD   100
mut-G114S    -----------------------------------------------G----------------------------------------------------
wt-Er        -----------------------------------------------S----------------------------------------------------
mut-S114G    -----------------------------------------------S----------------------------------------------------

wt-Em        LLAANIKIFKSQGGALDTVAKKTVKVVVVGNPANTNALICSKYAPSIPKENFSCLTRLDQNRAQAQIAARLGIPNSAVSNVIIWGNHSSTQFPDVRHAKA   200
mut-G114S    -------------S--------------------------------------------------------------------------------------
wt-Er        -------------S--------------------------------------------------------------------------------------
mut-S114G    -------------G--------------------------------------------------------------------------------------

wt-Em        IVNGKTVPVPEAVKDDNYLKNEFLTTVQKRGGEVIKARKLSSAMSAAKATADHMRDWWFGTQGDNWVSMGVFSTGAYGIPEGIMYSYPVRIANKQWKIVE   300
mut-G114S    ----------------------------------------------------------------------------------------------------
wt-Er        ----------------------------------------------------------------------------------------------------
mut-S114G    ----------------------------------------------------------------------------------------------------

wt-Em        GLEINDFAREKMDLTAKELCEERAMAEEICNQ   332
mut-G114S    --------------------------------
wt-Er        --------------------------------
mut-S114G    --------------------------------
```

Fig. S2. Deduced amino acid sequence alignments of cMDH from two 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).

```
wt-Em           MSEPVKVCVTGAAGQIAYSLLYSIAKGDVFGSDQPIILTLLDIAPMEGVLNGVVMELMDCAMPLLKEVIPTCDEKVAFKDIDAAFLVGAMPRREGMERKD   100
mut-P4A         ---A------------------------------------------------------------------------------------------------
wt-Lk           ---A------------------------------------------------------------------------------------------------
mut-A4P         ---P------------------------------------------------------------------------------------------------

wt-Em           LLAANIKIFKSQGGALDTVAKKTVKVVVVGNPANTNALICSKYAPSIPKENFSCLTRLDQNRAQAQIAARLGIPNSAVSNVIIWGNHSSTQFPDVRHAKA   200
mut-P4A         -------------G--------------------------------------------------------------------------------------
wt-Lk           -------------A--------------------------------------------------------------------------------------
mut-A4P         -------------A--------------------------------------------------------------------------------------

wt-Em           IVNGKTVPVPEAVKDDNYLKNEFLTTVQKRGGEVIKARKLSSAMSAAKATADHMRDWWFGTQGDNWVSMGVFSTGAYGIPEGIMYSYPVRIANKQWKIVE   300
mut-P4A         ----------------------------------------------------------------------------------------------------
wt-Lk           ----------------------------------------------------------------------------------------------------
mut-A4P         ----------------------------------------------------------------------------------------------------

wt-Em           GLEINDFAREKMDLTAKELCEERAMAEEICNQ   332
mut-P4A         --------------------------------
wt-Lk           --------------------------------
mut-A4P         --------------------------------
```

Fig. S3. Deduced amino acid sequence alignments of cMDH of two wild types, *Echinolittorina malaccana* (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).

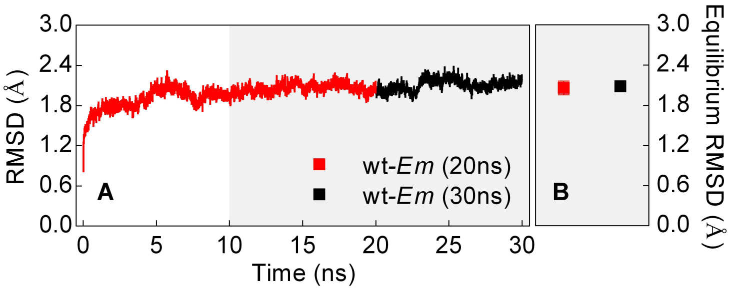

Fig. S4. The rmsd of backbone atom positions for the wild type cMDH, *Echinolittorina malaccana* (wt-*Em*) at a simulation temperature of 57 ℃. (*A*): Simulations for 20 ns and 30 ns, respectively. (*B*): The corresponding equilibrium rmsd values (10–20 ns and 10–30 ns, respectively) ($n = 3$), data are expressed as means ± SEM and no significant difference is found between the equilibrium values of the two timeframe simulations ($P = 0.851$).

Table S1. Overall and region-specific variation in different categories of amino acids of 26 cMDH orthologs of marine molluscs as a function of enzyme stability. Amino acid categories were: charged (DEKR), weakly hydrophobic/polar (AGNQSTHY), and hydrophobic (LVWIFMPC)

	Overall	Surface	Intermediate	Core
Charged residues				
Equation	Y = -7.449*X + 22.12	Y = -7,139*X + 17.31	Y = -0.204*X + 4.73	Y = -0.156*X + 0.09
R^2	0.541	0.397	<0.001	0.009
P value	<0.001	<0.001	0.906	0.646
Weakly hydrophobic/polar residues				
Equation	Y = 6.283*X + 40.11	Y = 6.454*X + 19.97	Y = -0.846*X + 16.30	Y = 0.676*X + 3.85
R^2	0.428	0.072	0.002	0.009
P value	<0.001	0.185	0.834	0.641
Hydrophobic residues				
Equation	Y = 1.168*X + 37.78	Y = -1.091*X + 11.01	Y = -0.083*X + 21.33	Y = 2.122*X + 5.383
R^2	0.035	0.021	<0.001	0.037
P value	0.362	0.482	0.969	0.348
Proline				
Equation	Y = 2.293*X + 4.29	Y = 1.932*X + 2.99	Y = 0.249*X + 1.17	Y = 0.112*X + 0.13
R^2	0.182	0.183	0.007	0.004
P value	0.030	0.029	0.679	0.758
Glycine				
Equation	Y = 2.073*X + 6.54	Y = 2.241*X + 3.99	Y = -0.248*X + 1.82	Y = 0.181*X + 0.72
R^2	0.186	0.209	0.002	0.007
P value	0.028	0.019	0.834	0.694

Table S2. The rmsd of backbone atom positions and the rms fluctuation (rmsf) for individual residues over the equilibration state (10–20 ns) determined by molecular dynamics simulations for cMDHs at the simulation temperature of 57 ℃ (mean ± SEM). Two 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) (*n* = 10); two wild types, wt-*Em* and *Littorina keenae* (wt-*Lk*), and two mutants of 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)

Species	rmsd	rmsf
wt-*Em*	2.117 ± 0.049	1.075 ± 0.017
mut-G114S	1.986 ± 0.059	1.019 ± 0.020
wt-*Er*	2.183 ± 0.089	1.066 ± 0.029
mut-S114G	2.665 ± 0.107	1.213 ± 0.033
wt-*Em*	2.110 ± 0.067	1.072 ± 0.025
mut-P4A	2.104 ± 0.067	1.080 ± 0.025
wt-*Lk*	2.324 ± 0.098	1.151 ± 0.037
mut-A4P	2.017 ± 0.047	1.037 ± 0.009

Table S3. Statistical analyses of the averaged equilibrium (10–20 ns) rmsd of backbone atom positions and averaged equilibrium rms fluctuation (rmsf) for individual residues determined by molecular dynamics simulations for cMDHs at the simulation temperature of 57 ℃. Two 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) (n = 10); two 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) (n = 5). Turkey's multiple comparisons test with Prism 6.0 software (La Jolla, CA, USA) was used. Significance is denoted with asterisks: * for P value less than 0.05, ** for P values less than 0.01, *** for P value less than 0.001, and ns for not significant

Species

	wt-*Em*	mut-G114S	wt-*Er*	mut-S114G
wt-*Em*		ns	ns	** 0.003
mut-G114S	ns		ns	*** < 0.001
wt-*Er*	ns	ns		** 0.001
mut-S114G	*** < 0.001	*** < 0.001	*** < 0.001	
	wt-*Em*	mut-P4A	wt-*Lk*	mut-A4P
wt-*Em*		ns	ns	ns
mut-P4A	ns		ns	ns
wt-*Lk*	ns	ns		* 0.029
mut-A4P	ns	ns	* 0.038	

One-way ANOVA of rmsd value

One-way ANOVA of rmsf value

Table S4. Statistical analyses of the temperature-dependence of the Michaelis-Menten constant for the cofactor (reduced nicotinamide adenine dinucleotide) (K_M^{NADH}). Three wild types *Echinolittorina malaccana* (wt-*Em*), *E. radiata* (wt-*Er*), and *Littorina keenae* (wt-*Lk*), and four mutants, mut-G114S (glycine mutated to serine at site 114 for the *E. malaccana* ortholog), mut-S114G (serine mutated to glycine for the *E. radiata* ortholog), 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 independent sample t-test with Prism 6.0 software (La Jolla, CA, USA) was used. Significance is denoted with asterisks: * for P value less than 0.05, ** for P values less than 0.01, *** for P value less than 0.001, and ns for not significant

	Temperature		Statistical analyses			
			wt-*Em*	mut-G114S	wt-*Er*	mut-S114G
K_M^{NADH}	20 ℃	wt-*Em*	-	ns	ns	ns
		mut-G114S	-	-	ns	ns
		wt-*Er*	-	-	-	ns
		mut-S114G	-	-	-	-
	25 ℃	wt-*Em*	-	* 0.033	ns	** 0.002
		mut-G114S	-	-	ns	ns
		wt-*Er*	-	-	-	ns
		mut-S114G	-	-	-	-
	30 ℃	wt-*Em*	-	* 0.029	* 0.034	** 0.002
		mut-G114S	-	-	ns	** 0.007
		wt-*Er*	-	-	-	* 0.049
		mut-S114G	-	-	-	-
	35 ℃	wt-*Em*	-	*** < 0.001	*** < 0.001	*** < 0.001
		mut-G114S	-	-	ns	*** < 0.001
		wt-*Er*	-	-	-	* 0.010
		mut-S114G	-	-	-	-
	40 ℃	wt-*Em*	-	*** < 0.001	*** < 0.001	*** < 0.001
		mut-G114S	-	-	ns	** 0.007
		wt-*Er*	-	-	-	ns
		mut-S114G	-	-	-	-
Linear regression (slope)		wt-*Em*	-	* 0.017	** 0.002	*** < 0.001
		mut-G114S	-	-	ns	ns
		wt-*Er*	-	-	-	ns
		mut-S114G	-	-	-	-
			wt-*Em*	mut-P4A	wt-*Lk*	mut-A4P
K_M^{NADH}	20 ℃	wt-*Em*	-	* 0.047	ns	ns
		mut-P4A	-	-	* 0.037	ns
		wt-*Lk*	-	-	-	ns
		mut-A4P	-	-	-	-
	25 ℃	wt-*Em*	-	*** < 0.001	* 0.047	** 0.008
		mut-P4A	-	-	ns	ns
		wt-*Lk*	-	-	-	ns
		mut-A4P	-	-	-	-
	30 ℃	wt-*Em*	-	** 0.002	ns	* 0.018
		mut-P4A	-	-	* 0.018	* 0.025
		wt-*Lk*	-	-	-	ns
		mut-A4P	-	-	-	-
	35 ℃	wt-*Em*	-	*** < 0.001	** 0.006	** 0.002
		mut-P4A	-	-	* 0.023	ns
		wt-*Lk*	-	-	-	ns
		mut-A4P	-	-	-	-
	40 ℃	wt-*Em*	-	*** < 0.001	** 0.004	*** < 0.001
		mut-P4A	-	-	** 0.003	* 0.032
		wt-*Lk*	-	-	-	ns
		mut-A4P	-	-	-	-
Linear regression (slope)		wt-*Em*	-	* 0.018	ns	ns
		mut-P4A	-	-	ns	ns
		wt-*Lk*	-	-	-	ns
		mut-A4P	-	-	-	-

Table S5. Primers designed for cytosolic malate dehydrogenase (cMDH) cDNAs

Primer ID	Description	Sequences (5’–3’)
MDF1	Partial sequence primer for *Laternula elliptica or Adamussium colbecki*	CTGCKGGYCARATYGCYTAYT
MDR1		CTGCWGACATRGCACTRGAYA
Le-5race-1	RACE primer for *L. elliptica*	GATTACGCCAAGCTTTAACATCTGGATACTGGGTTGAGGA
Le-5race-2		GATTACGCCAAGCTTCATGGCTCCAACCAGAATGGCTAC
Le-3race-1		GATTACGCCAAGCTTCTTGCTAAGGGAGATGTATTTGGAC
Le-3race-2		GATTACGCCAAGCTTGATTACGCAGTTGGCTTCAAGGATG
Ac-5race-1	RACE primer of *A. colbecki*	GATTACGCCAAGCTTCTTGACCGTGGCGTAGGAGACAT
Ac-5race-2		GATTACGCCAAGCTTTAGGCACCACATTTCTGACAAGG
Ac-3race-1		GATTACGCCAAGCTTAGGCACAGGTTGCTAATAAGTTGG
Ac-3race-2		GATTACGCCAAGCTTTTATCAGCACTGTCCAGAAGCGAGG
E-F1	Full-length primer with restriction site (underlined) for *Echinolittorina malaccana or E. radiata*	CCGGAATTCATGAGTGAACCAGTGAAAG
E-F2		CCGGAATTCATGAGTGAACCAGTGAAAGTATGT
E-R1		CCGCTCGAGTCACTGGTTGCAGATTTCTTCG
E-R2		CCGCTCGAGTCACTGGTTGCAGATTTCTTCG
Em-G114S-F1	114# mutagenesis primer for *E. malaccana or E. radiata* to construct the mut-G114S and mut-S114G	TTCAAGTCCCAGGGCTCCGCCCTTGACACTGT
Em-G114S-F2		CAAGTCCCAGGGCTCCGCCCTTGACACT
Er-S114G-F1		TTCAAGTCCCAGGGCGGTGCTCTTGACACCGT
Er-S114G-F2		CAAGTCCCAGGGCGGTGCTCTTGACACC
E-114-R1		GATCTTGATGTTGGCTGCCAGCAGGTCCTTGC
E-114-R2		AAGATCTTGATGTTGGCTGCCAGCAGGTCCT
Em-P4A-F1	4# mutagenesis primer for *E. malaccana* to construct the mut-P4A	GTAACTGGTGCTGCTGGC
Em-P4A-R1		GCATACTTTCACTGCTTCAC
Em-P4A-F2		TGGTGCTGCTGGCCAGATTGCT
Em-P4A-R2		GTTACGCATACTTTCACTGCTTCACTCAT
G114A-F1	4# mutagenesis primer for *Littorina keenae* to construct the wt-*Lk* and mut-A4P	TGACACTGTGGCCAAGAAGACCGTCAAGGT
G114A-R1		AGGGCGGCGCCCTGGGACTTGAAGAT
G114A-F2		ATCTTCAAGTCCCAGGGCGCCGC
G114A-R2		CTTGATGTTGGCTGCCAGCAGGTCCTT

Table S6. GenBank accession Numbers for 26 orthologs of molluscan cytosolic malate dehydrogenase (cMDH) sequences

Species	Accession Number
Echinolittorina radiata	MF774407
E. malaccana	MF774406
Littorina keenae	MG256005
L. scutulata	MG256006
Nipponacmea fuscoviridis	MF774412
N. radula	MF774413
Lottia gigantea	EU863455
L. austrodigitalis	EU863454
L. scabra	EU863457
L. digitalis	EU863452
L. scutum	EU863456
L. pelta	EU863453
Chlorostoma rugosa	MG190819
C. funebralis	MG190818
C. brunnea	MG190820
C. montereyi	MG190821
Nerita yoldii	MF774409
N. albicilla	MF774408
Haliotis rufescens	MG256004
H. diversicolor	MF774410
H. discus	MF774411
Mytilus trossulus	DQ149969
M. californianus	DQ149971
M. galloprovincialis	DQ149970
Laternula elliptica	MH924592
Adamussium colbecki	MH924591