Biophysical Journal, Volume 98

Supporting Material

Thermal Adaptation of Viruses and Bacteria

Peiqiu Chen and Eugene Shakhnovich

Supplementary Information

A. Details of simulations.

The genotype of the daughter organism can be presented as

$$\left\{\Delta\vec{G}\right\}_{daughter} = \left\{\Delta\vec{G}\right\}_{parent} + \left\{\Delta\Delta\vec{G}\right\}$$

where $\left\{\Delta\Delta\vec{G}\right\} = \left(\Delta\Delta G_{i1}, \Delta\Delta G_{i2} .... \Delta\Delta G_{i_s}\right)$ describes changes of stabilities upon a replication event which resulted in s mutations in proteins $\left(i_1, i_2 .... i_{i_s}\right)$. For semi-conservative replication, mutations might occur in both the parent copy and the descendent copy. For conservative replication, mutations would then occur in the descendent copy only. We generate the number of mutations s at each replication in a daughter organism, according to a Poisson distribution, and the parameter of the Poisson distribution $m_{organism}$ is the average number of mutations per genome per replication, for this particular species. The mutation rate for each gene in each copy is then $m_{gene} = m_{organism} / \Gamma$. After selecting s - the total number of proteins to be mutated at a given replication event - we decide which proteins to mutate by selecting the set $\left(i_1, i_2 ... i_s\right)$ at random. When a mutation occurs in a protein, the protein's sequence would be physically changed. In this study we do not consider protein sequences explicitly. Rather we posit that free energy of the mutant protein would have a different folding free energy (1), and the free energy difference $\Delta\Delta G_i$ between wild-type and mutant protein is a random value drawn from a distribution based on statistics of free energy changes collected in multiple protein engineering experiments (2). To this end, we determine the statistics of changes of protein stability upon mutations from the ProTherm database (1) as it was done in our earlier study (2). This database contains information on more than three thousand point mutations, derived from most currently performed point mutation experiments. The statistics show that protein folding stability change due to point mutation roughly forms a Gaussian distribution, where the mean is 1kcal/mol and the standard deviation is 1.7kcal/mol (2). Therefore, when a mutation occurs, we alter the protein stability by an amount drawn from this Gaussian distribution. We assume that statistics of changes of protein stability $\Delta\Delta G$ does not depend on stability ΔG itself. A similar assumption was made in (3). In a recent study (4) we confirmed the validity of this assumption. The mutant daughter organism will therefore have an altered fitness value (derived from (Eq. (2)), due to altered stability of some of its proteins.

B. Derivation of Various Coefficients

According to the analysis in the main text, for instant increase of temperature to $T + \delta T$, for $-L + \Delta S_R k_B \delta T < \tilde{G}_i < \Delta S_R k_B \delta T$, the relationship between free energy change of an individual protein and change in free energy distribution can be presented as:

$$\begin{cases} \tilde{G}_i = G_i + \Delta S_R k_B \delta T \\ p(\tilde{G}_i) = C_0(-e^{\frac{\tilde{G}_i - \overline{\Delta S} k_B \delta T}{D}} Sin[\pi \frac{\tilde{G}_i - \Delta S_R k_B \delta T}{L}]) \end{cases}$$

The birth rate at temperature $T + \delta T$ can therefore be expressed as:

$$< \ln b(T + \delta T) >= \ln b_0 - \frac{H^{\#}}{k_B(T + \delta T)} - \Gamma \int_{-L+\overline{\Delta S} k_B \delta T}^{\overline{\Delta S} k_B \delta T} \ln(1 + e^{\frac{\tilde{G}_i}{k_B(T+\delta T)}}) p(\tilde{G}_i) d\tilde{G}_i$$

The first term in the right hand side is constant with respect to temperature. The second term, $-\frac{H^{\#}}{k_B(T+\delta T)}$ is the metabolic reaction barrier term, and it increases upon the increase of temperature. The third term can be evaluated by perturbation, and it's behavior upon the change of temperature can be studied in this way as well.

From the integration term above, denote $y = \tilde{G}_i - \Delta S_R k_B \delta T$, then the integration part changes to:

$$\int_{-L+\overline{\Delta S} k_B \delta T}^{\overline{\Delta S} k_B \delta T} \ln(1 + e^{\frac{\tilde{G}_i}{k_B(T+\delta T)}}) p(\tilde{G}_i) d\tilde{G}_i = \int_{-L}^{0} \ln(1 + e^{\frac{y + \overline{\Delta S} k_B \delta T}{k_B(T+\delta T)}})(-C_0)(e^{\frac{y}{\Psi}} Sin[\frac{\pi y}{L}]) dy$$

Therefore, the birth rate at temperature $\delta T + T$ can be expressed as:

$$< \ln b(T + \delta T) >= \ln b_0 - \frac{H^{\#}}{k_B(T + \delta T)} - \Gamma \int_{-L}^{0} \ln(1 + e^{\frac{G + \Delta S_R k_B \delta T}{k_B(T+\delta T)}}) p(G) dG \qquad \text{(S1)}$$

Denote the third term as:

$$I(T + \delta T) = -\Gamma \int_{-L}^{0} \ln(1 + e^{\frac{G + \Delta S_R k_B \delta T}{k_B(T+\delta T)}}) p(G) dG$$

This term represents the average logarithm concentration of the folded protein at temperature $T + \delta T$. $\ln(1 + e^{\frac{G + \Delta S_R k_B \delta T}{k_B(T+\delta T)}})$ can be expanded as:

$$\ln(1 + e^{\frac{G + \Delta S_R k_B \delta T}{k_B(T+\delta T)}}) = -\frac{\delta T}{k_B T^2} f_1(T) + \frac{1}{2}(\frac{\delta T}{k_B T^2})^2 f_2(T),$$

where:

$$\begin{cases} f_1(T) = \dfrac{e^{\frac{G}{k_B T}}}{(1+e^{\frac{G}{k_B T}})}(G - \Delta S_R k_B T) \\ f_2(T) = \dfrac{e^{\frac{G}{k_B T}}}{(1+e^{\frac{G}{k_B T}})^2}(G - \Delta S_R k_B T)(G - \Delta S_R k_B T + 2k_B T * (1+e^{\frac{G}{k_B T}})) \end{cases} \quad \text{(S2)}$$

We can express the log ratio of folded protein at $T + \delta T$ as:

$$\frac{I(T+\delta T)}{\Gamma} = \frac{\delta T}{k_B T^2}\int_{-L}^{0} f_1(T)p(G)dG - \frac{1}{2}(\frac{\delta T}{k_B T^2})^2 \int_{-L}^{0} f_2(T)p(G)dG \quad \text{(S3)}$$

It is possible to evaluate this integral fully analytically for both the first and the second order term, however the complete integration result is lengthy. However, since for mesophiles, T~300K, we have $L >> k_B T$, $\Delta S_R k_B * T >> G$. Therefore we can approximate (S2) as:

$$\begin{cases} f_1(T) = -\Delta S_R k_B T e^{\frac{G}{k_B T}} \\ f_2(T) = (-\Delta S_R k_B T)^2 e^{\frac{G}{k_B T}} \end{cases} \quad \text{(S4)}$$

This is a significant simplification. Evaluating (S3) is straightforward after employing the simplified functional form of (S4), by noticing $e^{-\frac{L}{k_B T}} -> 0, e^{-\frac{L}{\Psi}} -> 0$, the integration result can be further approximately expressed as:

$$\int_{-L}^{0} f_1(T)p(G)dG = \frac{\Delta S_R k_B T * (k_B T)^2}{(\Psi^2 + (k_B T)^2)}, \qquad \int_{-L}^{0} f_2(T)p(G)dG = -\frac{(\Delta S_R k_B T)^2 * (k_B T)^2}{(\Psi^2 + (k_B T)^2)}$$

(S5)

According to the analysis above, we can express the logarithmic ratio of birth rate as:

$$< \ln(\frac{b(T+\delta T)}{b(T)}) >= \delta T(\Gamma C_1 + C_2) + \delta T^2(\Gamma C_3 + C_4),$$

where the coefficients are

$$\begin{cases} C_1 \sim -\dfrac{\Delta S_R k_B * k_B T}{(\Psi^2 + (k_B T)^2)} \\ C_2 = \dfrac{H}{k_B T^2} \\ C_3 \sim -\dfrac{(\Delta S_R k_B)^2}{2(\Psi^2 + (k_B T)^2)} \\ C_4 = -\dfrac{H}{k_B T^3} \end{cases} . \qquad \text{(S6)}$$

These values can then be used in the analysis of how the metabolic reaction free energy barrier and rate determining gene number in each species affect their thermal response behavior, as explained in the main text.

Supplementary Figure 1

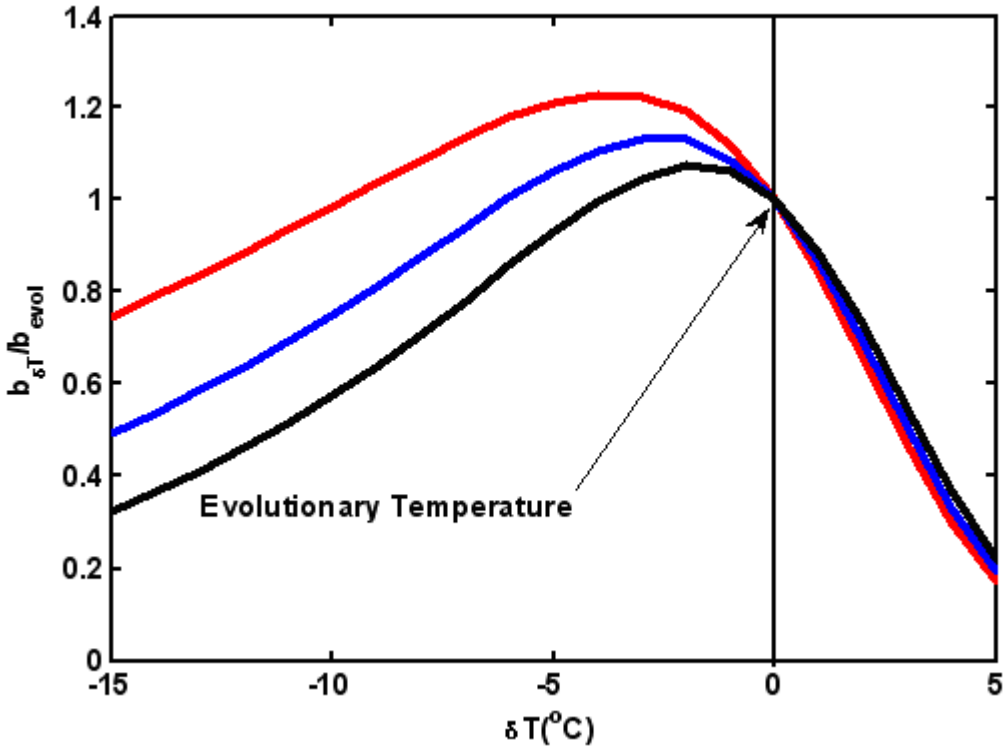

Figure Caption. The dependence of the growth rate on temperature for organisms whose genome sizes are at the optimal RDG number Γ_C. Red line (light grey in print); $H^{\#}$ = 10, blue line (dark grey in print); $H^{\#}$ = 15, Black line; $H^{\#}$ = 20. T = 37°C and Γ = 20 for all three species.

C. Thermal response curve fits for all 35 Datasets

Here we provide a table for all 35 mesophilic bacteria whose thermal response has been studied (5). The correlation between experimental fit and our theoretical prediction are from 90% to 99%.

	H	Γ	**correlation**
lmono 1	7.6483	24	97.88%
lmono2	6.608	22	96.08%
lmono 3	7.3372	24	96.68%
lmono 4	7.6	38	97.75%
lmono 5	7.2874	35	97.48%
g punicea	7.2387	11	96.43%
galidibacter	4.8427	11	89.96%
shewanella	7.3522	37	96.71%
shewanella	7.9265	22	99.33%
shewanella	6.0716	20	97.88%
a. hydrophila	8.5582	14	99.16%
l mono scott	8.4017	18	98.84%
e coli m23	10.5478	19	97.77%
ps florescence 1412	8.0594	34	98.76%
kleb oxy	8.1544	19	98.39%
p putida 1412	6.9578	36	97.57%
K120-6	11.8016	32	93.07%
K118-4	11.1957	46	98.48%
BC-29 (Exp. 3)	9.2917	32	96.92%
BC-29 (Exp. 2)	9.4541	42	98.75%
BC-29 (Exp. 1)	11.102	41	98.18%
BC-14 (Exp. 5)	9.4761	13	97.53%

BC-14 (Exp. 3)	10.0644	32	99.01%
BC-14 (Exp. 2)	7.2467	12	98.88%
BC-14 (Exp. 1)	10.729	31	98.73%
BC-14 (Exp. 3)	8.8185	18	97.26%
E.coli ONT H8 (R91) Mark Salter's thesis:	8.6747	27	98.18%
E.coli O126:H21(R10) Mark Salter's thesis:	7.6515	24	98.10%
E.coli NT (R31) Mark Salter's thesis:	8.5031	45	97.24%
O81:H- (R106)	8.8661	46	97.80%
O88:H- (R171)	8.8661	46	97.80%
O88:H- (R172)	8.7549	48	97.60%
O157:H-	9.0366	40	97.75%
O157:H7 (EH9)	8.8526	19	98.04%
O111:H-	8.6179	20	98.50%

1. Kumar, M. D., K. A. Bava, M. M. Gromiha, P. Prabakaran, K. Kitajima, H. Uedaira, and A. Sarai. 2006. ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions. Nucleic Acids Res 34:D204-206.
2. Zeldovich, K. B., P. Chen, and E. I. Shakhnovich. 2007. Protein stability imposes limits on organism complexity and speed of molecular evolution. Proc Natl Acad Sci U S A 104:16152-16157.
3. Bloom, J. D., A. Raval, and C. O. Wilke. 2007. Thermodynamics of neutral protein evolution. Genetics 175:255-266.
4. Chen, P., and E. I. Shakhnovich. 2009. Lethal mutagenesis in viruses and bacteria. Genetics 183:639-650.
5. Ratkowsky, D. A., J. Olley, and T. Ross. 2005. Unifying temperature effects on the growth rate of bacteria and the stability of globular proteins. J Theor Biol 233:351-362.