The effective temperature of mutations

Imre Derényi[*] and Gergely J. Szöllősi[†]

ELTE-MTA "Lendulet" Biophysics Research Group, Department of Biological
Physics, Eötvös University, Pázmány P. stny. 1A, H-1117 Budapest, Hungary

Biological macromolecules experience two seemingly very different types of noise acting on different time scales: i) point mutations corresponding to changes in molecular sequence and ii) thermal fluctuations. Examining the secondary structures of a large number of microRNA precursor sequences and model lattice proteins, we show that the effects of single point mutations are statistically indistinguishable from those of an increase in temperature by a few tens of Kelvins. The existence of such an effective mutational temperature establishes a quantitative connection between robustness to genetic (mutational) and environmental (thermal) perturbations.

Biological systems are robust [1], and robustness is considered to be a fundamental feature of complex evolvable systems [2]. Molecular phenotypes, such as stable protein folds and functional RNA structures, have provided fundamental insight into the origin and principles of robustness in biological systems [5–12]. In these systems perturbations on different time scales, i.e., rare changes in sequence and omnipresent thermal fluctuations, seem very different. Surprisingly, several computational [13–16] and experimental [17–21] studies suggest a qualitative similarity between the effects of mutations and temperature. Stable proteins are in general more tolerant to point mutations [14, 17], and in the case of RNA, the set of structures explored by thermal fluctuations are highly correlated with the minimum free energy structures of single point mutants[13, 16].

The correlation between the effects of point mutations and temperature is less surprising if we recognize that each degree of freedom of the molecule has an average thermal energy of $k_\mathrm{B}T/2 \approx 2.5/2$ kJ/mol (where k_B is the Boltzmann constant and $T \approx 300$ K is the absolute temperature) and the typical free energy change associated with a point mutation is also of the order of a $k_\mathrm{B}T$ (approximately 4 kJ/mol for a protein [22], and about twice this large for the breaking of a hydrogen bond in a nucleotide base pair). Despite this similarity in energies, for single instances of the system, i.e., individual copies of protein or RNA molecules, permanent changes in sequence are clearly different from ephemeral thermal kicks. However, the distinction between mutational effects and thermal fluctuations becomes less manifest in large populations and over longer time scales where many possible point mutations are explored as molecules are copied (transcribed and translated) repeatedly via mechanisms prone to errors. From this perspective, perturbations of the phenotype (e.g., the protein fold or RNA secondary structure) resulting from mutations can be expected to have similar effects to thermal perturbations: both jostle the system between states with energies that differ by only a few times the thermal energy scale.

Here we demonstrate that this qualitative analogy between mutational and thermal perturbations can be taken to a quantitative level, and the effect of point mutations is well described as an effective increase in temperature. For the original, so

---

called wild type (WT), sequence let the probability of the system being in any of its possible states (e.g., protein folds or RNA secondary structures) be denoted by

$$P_\mathrm{WT}(i,T) = \frac{1}{Z_\mathrm{WT}(T)} \exp\left(-\frac{G_\mathrm{WT}(i,T)}{k_\mathrm{B}T}\right) , \quad (1)$$

where $G_\mathrm{WT}(i,T)$ is the free energy of state i and $Z_\mathrm{WT}(T)$ is the partition function, and let the probability of the same state transformed to an effective temperature T_eff be defined as

$$P_\mathrm{eff}(i,T,T_\mathrm{eff}) = \frac{1}{Z_\mathrm{eff}(T,T_\mathrm{eff})} \exp\left(-\frac{G_\mathrm{WT}(i,T)}{k_\mathrm{B}T_\mathrm{eff}}\right) , \quad (2)$$

with normalization constant $Z_\mathrm{eff}(T,T_\mathrm{eff})$. Note that this transformation preserves the relative contributions of the enthalpic and entropic components of the free energy at the original temperature T. The mutation averaged probability of the system in state i can be defined as

$$P_\mathrm{mut}(i,T) = \frac{1}{N_\mathrm{mut}} \sum_{m=1}^{N_\mathrm{mut}} p_m(i,T) , \quad (3)$$

where the summation index m runs over all possible point mutations (of total number N_mut) and $p_m(i,T)$ denotes the probability of state i for mutation m.

Examining the secondary structures of a large number of microRNA (miRNA) precursor sequences and model lattice proteins, we demonstrate that in most cases there exists a well defined effective temperature T_eff^*, for which the transformed probability distribution of the wild type approximates the average probability distribution of single point mutants with unanticipated precision:

$$P_\mathrm{mut}(i,T) \approx P_\mathrm{eff}(i,T,T_\mathrm{eff}^*) . \quad (4)$$

In other words, we show that the effect of mutations are fully described by considering only the free energies $G_\mathrm{WT}(i,T)$ of the secondary structures i of the single *wild type* sequence and a single effective temperature T_eff^*.

We used the 23766 miRNA precursor sequences of miRBase version 9.0 [23] upto the length of 250 nucleotides. For each of these WT sequences we (i) generated all of its single point mutants, and then (ii) determined the equilibrium probability distributions (at $T = 300$ K) of the secondary structures of the WT and the mutant sequences, by using the
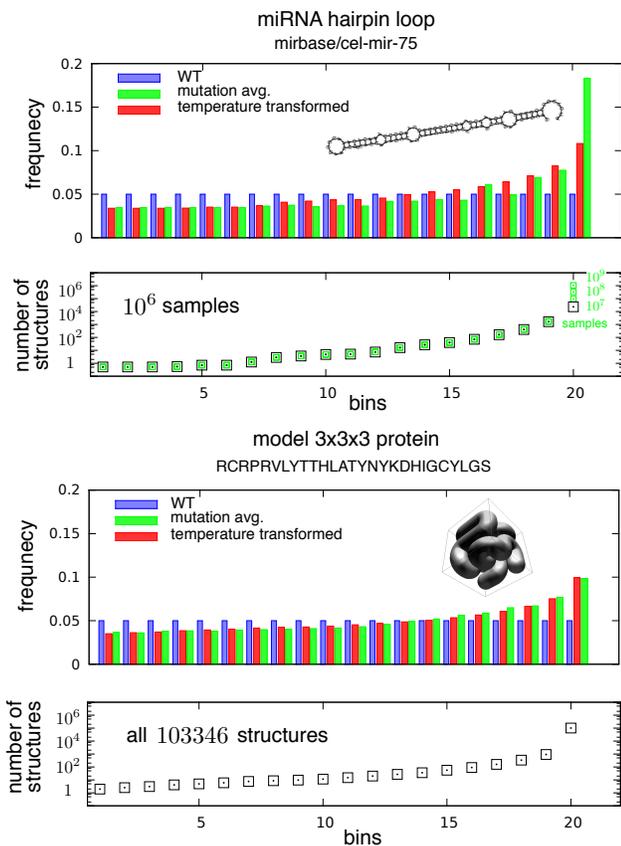
FIG. 1. **The effects of single point mutations and an increased effective temperature.** For a typical miRNA and lattice protein the structures were grouped into 20 bins in such a way that each bin accommodated exactly $1/20$th of the WT equilibrium probability distribution (blue bars). The number of structures for each bin is shown in the bottom panels. For the miRNA sequence smaller green squares show the result of sampling additional structures up to a total of 10^7, 10^8 and 10^9 samples. The histogram of the mutation averaged probability distribution (green bars) is tilted to the right, indicating that the mutations tend to destabilize the most stable structures. Raising the temperature has a similar effect, as demonstrated by the histogram of the WT distribution transformed to the optimal effective temperature T_{eff}^* defined in the main text (red bars).

computational tools developed by Hofacker et al. [24] and sampling 10^6 structures for each sequence. To establish the generality of our results we also analyzed a markedly different system, the model of $3 \times 3 \times 3$ compact lattice proteins [7]. We first randomly selected 12000 sequences, and designated them as the WT. Subsequently, similarly to the miRNA precursor sequences, for each of these WT sequences we (i) generated all of its single point mutants, and (ii) determined the equilibrium probability distributions of all 103346 possible 3D structures of the WT and the mutant sequences.

Two typical examples, a miRNA and a lattice protein, are shown in Fig. 1. Since depicting the probability distribution of a large number of secondary structures is not feasible, to obtain an easier to visualize representation, we grouped the structures into 20 bins in such a way that each bin accommo-

dated exactly $1/20$th of the WT equilibrium probability distribution (P_{WT}). The bins were filled from left to right with structures in decreasing order of their WT probability (starting with the most probable one, i.e., with the structure having the lowest free energy G_{WT}). To ensure that each bin contained equal probability, structures at bin boundaries were split among neighboring bins. The (generally non-integer) number of structures belonging to each bin is shown in the panels below the histograms. Examining Fig. 1 we can see that the histogram of the WT equilibrium distribution (P_{WT}, blue bars) is uniform by construction. The histogram of the mutation averaged probability distribution (P_{mut}, green bars), however, deviates from uniformity and is tilted to the right, indicating that the mutations tend to destabilize the most stable structures. Raising the temperature is expected to have a similar effect. Indeed, choosing the optimal effective temperature (T_{eff}^*), for which the Euclidean distance between the temperature transformed and mutation averaged distributions,

$$d(T_{\text{eff}}) = \sqrt{\sum_i \left[P_{\text{eff}}(i, T, T_{\text{eff}}) - P_{\text{mut}}(i, T) \right]^2} , \quad (5)$$

is minimal, the histogram of the transformed distribution (P_{eff}, red bars) approximates that of the mutation averaged distributions remarkably well. The only significant discrepancy (mostly observed for miRNAs) occurs at the rightmost bin, which contains rarely visited, high energy, non-native structures. Among the many possible statistical distances we chose the Euclidean distance because it is relatively insensitive to these statistically insignificant cases.

In the leftmost panels of Fig. 2 the Euclidean distance $d(T_{\text{eff}})$ (solid lines) is plotted as a function of the effective temperature T_{eff} for both examples from Fig. 1. If there existed a T_{eff} at which $d(T_{\text{eff}})$ were exactly zero, then the curve would have a "V"-shaped bottom. In reality, however, this is never the case, and the $d(T_{\text{eff}})$ curve has a non-zero parabolic minimum at T_{eff}^* (red arrow in the top left panel of Fig. 2), with second derivative $d''(T_{\text{eff}}^*)$. The half width (green interval in the top left panel of Fig. 2) of the fitting parabola (dashed lines) between the points where the parabola touches the two tangentials (dotted lines) that form a "V" as they intersect the horizontal axis at T_{eff}^*,

$$\delta T_{\text{eff}}^* = \sqrt{\frac{2d(T_{\text{eff}}^*)}{d''(T_{\text{eff}}^*)}} , \quad (6)$$

can be used to estimate the uncertainty in determining the effective temperature.

The middle and right panels of Fig. 2 show the distributions of the optimal effective temperatures T_{eff}^* and their uncertainties δT_{eff}^*, respectively, for all studied miRNAs (top) and model proteins (bottom). The typical uncertainty ($\sim 0.02T$ for the proteins and $\sim 0.03T$ for the miRNAs) is considerably smaller than both the width of the corresponding T_{eff}^* distribution and the typical value of T_{eff}^*, indicating the existence of a well defined optimal effective temperature for nearly all miRNAs and proteins.

For the miRNAs the length dependence of T_{eff}^* can also be examined (see Fig. 3, where the length L is defined as the
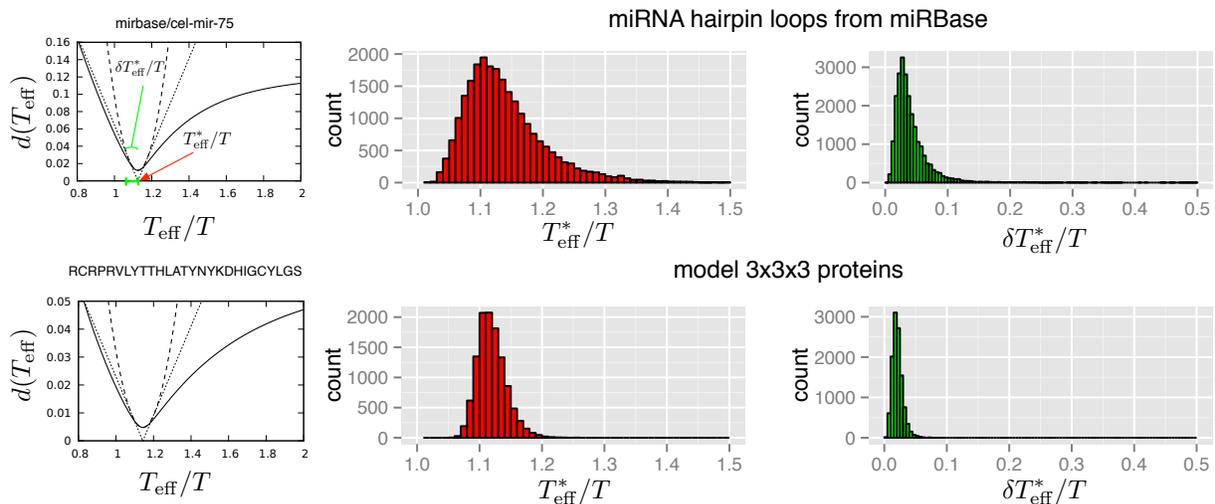
FIG. 2. **The distribution of the optimal effective temperature for miRNAs and lattice proteins.** In the leftmost panels the Euclidean distance $d(T_{\mathrm{eff}})$ (solid lines) is plotted as a function of the effective temperature T_{eff} for both examples from Fig. 1. The middle panels show the distributions of the optimal effective temperatures T_{eff}^* (red arrow in top left panel), while the rightmost panels show the distribution of uncertainties $\delta T_{\mathrm{eff}}^*$ (green interval in top left panel).
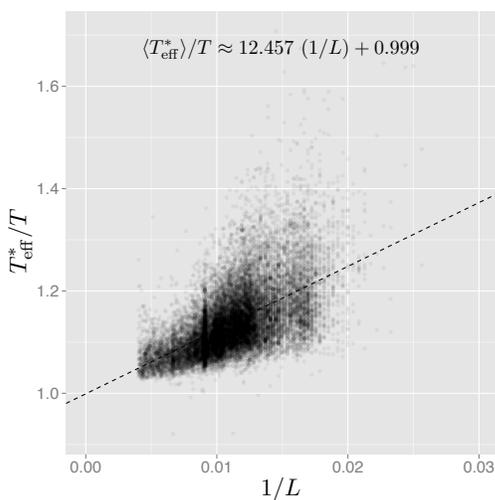

FIG. 3. **Relationship between sequence length and optimal effective temperature.** The optimal effective temperature is correlated with the reciprocal of the sequence length for the 23766 miRNA precursor sequences considered (Person's $\rho = 0.582$, with $p < 2.2 \times 10^{-16}$). Together with the intersection of the vertical axis at $T_{\mathrm{eff}}^*/T \approx 1$, this is consistent with the typical effect of a point mutation on the most stable structures being the breaking of a few hydrogen bonds and a corresponding increase of the free energy by about $12k_{\mathrm{B}}T$.

number of nucleotides). The data show that T_{eff}^* is roughly proportional to $1/L$, with a coefficient of about $12T$. This is not surprising if we realize that the most probable effect of a point mutation on the most stable structures is the breaking of a few hydrogen bonds and the concomitant increase of the free energy by about $12k_{\mathrm{B}}T$. Distributing this increase uniformly among the L nucleotides as a result of averaging over point mutations is analogous to raising the temperature by about $12T/L$.

A similar length dependence of T_{eff}^* is expected for the protein model (and any other systems), as well. Our data for $L = 27$ amino acids suggest a slope of about $3T$, which is consistent with an energy cost of about $3k_{\mathrm{B}}T$ of changing an amino acid in the most stable structures.

The diversity of proteins and nucleic acids we find in the living world are the result of evolution; their properties are determined by the laws of physics and chemistry [25]. To decipher the interplay of these two kinds of causality we have to understand the relationship between molecular sequence and function. The existence of a well defined mutational temperature demonstrates a general property of this relationship: the biophysics of RNA and protein structures imposes a statistical equivalence between the effects of mutational and thermal perturbations. In other words, the biological noise introduced by point mutations is quantitatively analogous to the physical noise generated by thermal fluctuations.

In an evolutionary context this implies that selection for robustness against either of these will produce, as a correlated by product, robustness against the other. Our result suggests an explicit model of how maintaining stability, a major constraint in the evolution of biological macromolecules, leads to stability against point mutations (both at the short time scales of molecular replication and the long time scales of organism reproduction), and also facilitates opportunities for molecular innovation by allowing increased neutral variation [17, 26].

[1] A. Wagner, *Robustness and evolvability in living systems* (Princeton University Press, Princeton, N.J., 2005).

[2] H. Kitano, Nat Rev Genet **5**, 826 (2004).

[3] R. E. Lenski, J. E. Barrick, and C. Ofria, PLoS Biol **4**, e428 (2006).

[4] B. P. Cusack, P. F. Arndt, L. Duret, and H. Roest Crollius, PLoS Genet **7**, e1002276 (2011).

[5] J. M. Smith, Nature **225**, 563 (1970).

[6] M. A. Huynen, P. F. Stadler, and W. Fontana, Proc Natl Acad Sci U S A **93**, 397 (1996).

[7] H. Li, R. Helling, C. Tang, and N. Wingreen, Science **273**, 666 (1996).

[8] E. van Nimwegen, J. P. Crutchfield, and M. Huynen, Proc Natl Acad Sci U S A **96**, 9716 (1999).

[9] K. Koelle, S. Cobey, B. Grenfell, and M. Pascual, Science **314**, 1898 (2006).

[10] G. J. Szöllősi and I. Derényi, Math Biosci **214**, 58 (2008).

[11] E. J. Hayden, E. Ferrada, and A. Wagner, Nature **474**, 92 (2011).

[12] A. Wagner, Trends Genet **27**, 397 (2011).

[13] L. W. Ancel and W. Fontana, J Exp Zool **288**, 242 (2000).

[14] J. D. Bloom, J. J. Silberg, C. O. Wilke, D. A. Drummond, C. Adami, and F. H. Arnold, Proc Natl Acad Sci U S A **102**, 606 (2005).

[15] A. Sakata, K. Hukushima, and K. Kaneko, Phys Rev Lett **102**, 148101 (2009).

[16] G. J. Szöllősi and I. Derényi, Mol Biol Evol **26**, 867 (2009).

[17] J. D. Bloom, S. T. Labthavikul, C. R. Otey, and F. H. Arnold, Proc Natl Acad Sci U S A **103**, 5869 (2006).

[18] P. Domingo-Calap, M. Pereira-Gómez, and R. Sanjuán, J Evol Biol **23**, 2453 (2010).

[19] A. S. Lauring, J. Frydman, and R. Andino, Nat Rev Microbiol **11**, 327 (2013).

[20] J. L. Payne and A. Wagner, Science **343**, 875 (2014).

[21] K. Kümpornsin, C. Modchang, A. Heinberg, E. H. Ekland, P. Jirawatcharadech, P. Chobson, N. Suwanakitti, S. Chaotheing, P. Wilairat, K. W. Deitsch, S. Kamchonwongpaisan, D. A. Fidock, L. A. Kirkman, Y. Yuthavong, and T. Chookajorn, Mol Biol Evol (2014), 10.1093/molbev/msu140.

[22] C. S. Wylie and E. I. Shakhnovich, Proc Natl Acad Sci U S A **108**, 9916 (2011).

[23] S. Griffiths-Jones, R. J. Grocock, S. Van Dongen, A. Bateman, and A. J. Enright, Nucleic acids research **34**, D140 (2006).

[24] I. Hofacker, W. Fontana, P. Stadler, S. Bonhoeffer, M. Tacker, and P. Schuster, Monatsh. Chem. **125**, 167 (1994).

[25] M. J. Harms and J. W. Thornton, Nat Rev Genet **14**, 559 (2013).

[26] A. Wagner, Proc Biol Sci **275**, 91 (2008).