

Recently it was indicated that fitness costs due to misfolded proteins are a determinant of evolutionary rate and selection originating in protein stability is a driving force of protein evolution. Here we examine protein evolution under the selection maintaining protein stability.

Protein fitness is a generic form of fitness costs due to misfolded proteins; $s = c \exp(dG / kT) (1 - \exp(ddG / kT))$, where s and dG are selective advantage and stability change of a mutant protein, dG is the folding free energy of the wild-type protein, and c is a parameter representing protein abundance and indispensability. The distribution of ddG is approximated to be a bi-Gaussian distribution, which represents structurally slightly- or highly-constrained sites. Also, the mean of the distribution is negatively proportional to dG .

The evolution of this gene has an equilibrium point dG_e , the range of which is consistent with observed values in the ProTherm database. The probability distribution of K_a/K_s , the ratio of nonsynonymous to synonymous substitution rate per site, over fixed mutants in the vicinity of the equilibrium shows that nearly neutral selection is predominant only in low-abundant, non-essential proteins of $dG_e > -2.5$ kcal/mol. In the other proteins, positive selection on stabilizing mutations is significant to maintain protein stability at equilibrium as well as random drift on slightly negative mutations, although the average $\langle K_a/K_s \rangle$ is less than 1. Slow evolutionary rates can be caused by both high protein abundance/indispensability and large effective population size, which produces positive shifts of ddG through decreasing dG_e and strong structural constraints, which directly make ddG more positive. Protein abundance/indispensability more affect evolutionary rate for less constrained proteins, and structural constraint for less abundant, less essential proteins. The effect of protein indispensability on evolutionary rate may be hidden by the variation of protein abundance and detected only in low-abundant proteins.