

MAR11-2010-003767

Abstract for an Invited Paper
for the MAR11 Meeting of
the American Physical Society

Insights into protein evolution landscapes from folding models

EUGENE KOONIN, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health

Off-lattice models of protein folding were employed to investigate the origins of the evolutionary rate distributions and fitness landscapes. For each robust folder, the network of sequences that share its native structure is identified. The fitness of a sequence is a simple function of the number of misfolded molecules produced to reach a characteristic protein abundance. Fixation probabilities of mutants are computed under a simple population dynamics model, and the fold-averaged evolution rate is computed using a Markov chain on the fold network. The distribution of the logarithm of the evolution rates exhibits a peak with a long tail on the low rate side and resembles the universal empirical distribution of the evolutionary rates more closely than either distribution resembles the log-normal distribution. We next addressed the question of the extent of determinism in protein evolution. Limited empirical studies suggest that the fitness landscapes of protein evolution are significantly smoother, or more additive, than random landscapes. However, widespread sign epistasis seems to restrict evolution to a small fraction of available trajectories, thus making the evolutionary process substantially deterministic. Access to complete fitness landscapes within the model framework enables exhaustive analysis of evolutionary trajectories. The model landscapes were compared to a continuum of artificial landscapes of varying smoothness. In maximally smooth, fully additive landscapes, evolution cannot be predicted because all paths are accessible. However, a small amount of noise can make most paths inaccessible while preserving the overall structure of the landscape. Although the model landscapes are almost additive, most paths are non-monotonic with respect to fitness, so evolutionary trajectories can be approximately predicted. Thus, protein folding physics seems to dictate the universal distribution of the evolutionary rates of protein-coding genes and the quasi-deterministic character of evolution.