

Distribution of the Local Thermodynamic Stability in the Complete Genome Sequence

Shu-Yun Le, Wei-min Liu, Jih-H. Chen, Jacob V. Maizel, Jr.,
National Institutes of Health

Local folding in mRNAs is closely associated with biological functions. In this study we reveal the whole distribution of local thermodynamic stability in the complete genome sequence of *Helicobacter pylori* strain 26695. Local thermodynamic stability is measured by two standards z-scores: significance score and stability score. To estimate the distribution of thermodynamic stability, a model based on the noncentral Student's t distribution has been developed. Significant patterns of those extremes that are either very highly stable or more unstable than expected by chance are assessed. Our results indicate that the highly stable and more statistically significant folding regions are predominantly in noncoding sequences, while the highly unstable folding regions are predominantly in the protein coding sequences. The observed difference across the complete genomic sequence is very statistically significant by a χ^2 -test. These extreme patterns may be useful in searching for target sequences for long-chain antisense RNA and for locating potential RNA functional elements involved in regulation of gene expression.