

Spectral Analysis of periodicities in bHLH amino acid sequences

Zhi Wang and William R. Atchley
Biomathematics and Bioinformatics Graduate Program,
North Carolina State University,
Raleigh, NC 27695, USA

Abstract The combination of classical periodogram method and advanced power spectral analysis can be used to detect and evaluate possible periodicity in amino acid sequences from multiple sequence alignments. The basic helix-loop-helix protein domain (bHLH) was chosen to explore the applicability of the low-resolution periodogram method and high-resolution power spectral analysis. Periodicity of amino acid variability was investigated using site-by-site Boltzmann-Shannon entropy values that measure the extent of amino acid variability. Preliminary statistical tests including stationary test, trend test and white noise tests were conducted. The periodogram revealed a repeated pattern 3.77 aa and power spectral analysis revealed a repeated pattern 3.59 aa, both close to the α -helix period 3.6 aa. It's the first time to report that the variability pattern of α -helix, measured by entropy values, has almost the same period as the secondary structure in a quantitative sense. To explain the origin of the period of entropy values, the amino acid codes were transformed to numerical values representing multiple factor scores of amino acid physiochemical attributes. This transformation permits detecting the periodicity patterns of more biologically meaningful components. The factor scores representing polarity/hydrophobicity is not white noise and has a period of 3.54 aa revealed by power spectral analysis. However, other factor scores representing molecular size, charge, PK, and other attributes have no significant periodic components and are accepted as white noise. Mathematical models describing the entropy and factor score patterns were fitted. It's found that the partition of the basic region and helices can produce more accurate period estimates, which justify the importance of examining the assumption of stationary. Both periodogram and power spectral analysis can promote our understanding of protein structure, function and evolution through detecting periodicity that may indicate secondary structure information.