Covariance Correlations from Genome-Wide Homology Sequence Analysis of DHFR

Vivian Cody, William L. Duax, Robert Huether, *Hauptman-Woodward Medical Research Institute*, Buffalo, NY 14203, USA. E-mail: Cody@hwi.buffalo.edu

To test the hypothesis that active site residue changes among dihydrofolate reductase (DHFR) influence binding specificity, the HSSP alignments of all protein sequences from the SWISS-PROT TrEMBL database that had 30% identity with DHFR were retrieved and resulted in a list of 298 gene products: 177 prokaryotic and 121 eukaryotic entries. Analysis of these profiles at the 70% identity level revealed: (1) 21 residues that are highly conserved in both kingdoms, (2) 13 additional residues whose frequency of occupancy achieves the 70% or greater level of sequence identity in eukaryote species only, (3) 14 sites in which a significant change in the dominant amino acid occupancy occurs between prokaryotic and eukaryotic species, and (4) the precise location of six inserts encompassing from one to seven residues that separate the two gene families. These results suggest that there has been an evolution from prokaryote to lower eukaryote to humans of an increasingly more specific ensemble of residues whose covariance correlates with functional specificity. A preference for ring-ring stacking involving Tyr33 and Tyr179 was noted in human DHFR. The usage profile at positions 33 and 179 respectively are: Y39, F23, H17% / F83, Y8% for eukaryotes and H31, Y3, F3% / F47, Y32% for prokaryotes. In the sequence of *Mycobacterium tuberculosis* DHFR these positions are H30 and L153. Supported in part by GM51670 (VC) and DK026546 (WLD).

Keywords: genome, dihydrofolate reductase, profile