Signatures of coevolution in protein superfamilies[1] SAROSH FA-
TAKIA, STEFANO COSTANZI, CARSON CHOW, Laboratory of Biological Mod-
eling, NIDDK, NIH — Protein superfamilies, such as G protein-coupled receptors,
consist of a large number of evolutionarily related proteins. A multiple sequence
alignment (MSA) from such a superfamily can help identify signatures of evolution.
A way to detect amino acid residues (AAs) that are important for the structure
and/or function of proteins is to identify a cohort of AAs that evolve in tandem
to compensate for mutations at any of those positions. To characterize such coevo-
lutionary patterns, a MSA of homologous sequences was used to identify pairs of
AA positions from the alignment having statistically significant mutual information
(MI). Pairs of such MSA positions that had high MI were represented as a graph to
show multiple associations. In that graph, the MSA positions represented vertices
whose edges were linked if position pairs had high MI. The vertices with high de-
gree were validated to be evolutionarily correlated positions that were important for
structure and/or function. Using subsets of more recently evolved proteins from the
diverse superfamily, most of those positions were determined to be under purifying
selection. This comparative genomic analysis may help infer protein structure and
coevolution in protein-protein interactions.

Sarosh Fatakia
LBM, NIDDK, NIH

Date submitted: 20 Nov 2009   Electronic form version 1.4