EXPLORING THE PROTEIN UNIVERSE FROM GENERAL PRINCIPLES

L. Apeltsin[1]

Biomedical Informatics, UCSF, 525 Nelson Rising Ln #514
San Francisco, CA 94158
[*]Email: Leonard.Apeltsin@ucsf.edu

T. E. Ferrin

Ferrin Group, Resource for Biocomputing, Visualization, and Informatics, 600 16th St Box M/S 2240
San Francisco, CA 94158
Email: tef@cgl.ucsf.edu

The Protein Universe encompasses all possible proteins, overlaid in a spatial formation. Computing the geometry of the Protein Universe will allow for large-scale prediction of structural and functional properties associated with completely uncharacterized proteins. The geometry is based on topological interpretations of how individual protein families relate to one another. However, in order to rely on these predictions we must first be sure that the geometry is valid. Currently, there exists no metric for quantifying the accuracy of any proposed geometrical mapping across protein space. Our goal is to generate this metric by developing algorithms that take as input a potential Protein Space Geometry (PSG), and output a set of predictions, that may then be tested computationally and experimentally. We have developed one such algorithm, Protein Space Trace (PST), based on a set of general principles defining the significance of a valid PSG. PST analyzes a set of sequences overlaid with the associated PSG, in order to elucidate out all the functionally significant residues for the proteins under analysis. In our study, we analyze the PSG of the Solute Carrier Transferase Superfamily, a set of transmembrane proteins that are difficult to crystallize. Running PST on this geometry generations functional residue predictions that overlap exceedingly well with experimentally determined mutagenesis data. This helps confirm the biological significance of the Solute Carrier Transferase PSG, allowing us with confidence to select the best possible candidate for crystallization from the data, the SLC22 family, which forms a hub within the topology of the PSG.

[1] Corresponding author.