

# National Academies Keck *Futures Initiative* Conference

## Mathematical Models in Signaling Systems - June 16-18, 2004

### ***Analysis of Network Architecture***

#### *Evolutionary Rules for Specifying Protein Folding and Function*

Rama Ranganathan, MD, PhD  
Associate Professor of Pharmacology  
The University of Texas Southwestern Medical Center at Dallas

#### ***Abstract:***

Classical studies show that for many proteins, the information required for specifying the tertiary structure is contained in the amino acid sequence. However, the potential complexity of this information is truly enormous, a problem that makes defining the rules for protein folding difficult through either computational or experimental methods. The recent expansion of the sequence database for many protein families offers new approaches to this fundamental problem. In recent work, we have reported a new method (that we call the statistical coupling analysis (SCA)) for estimating thermodynamic interactions between sites on proteins through statistical analysis of large and diverse multiple sequence alignments. Mutagenesis experiments demonstrate good agreement between the sequence-based parameters and thermodynamic values, suggesting that the SCA method is indeed a good reporter of energetic interactions in proteins. For diverse protein families, a global mapping of amino acid interactions through the SCA reveals a surprisingly simple general architecture: a sparse set of co-evolving residues form networks that connect distant functional surfaces through the protein core. The relative simplicity of this pattern suggests an interesting hypothesis: the small set of statistical rules derived from the SCA might be sufficient to specify a protein fold and its characteristic function.

To test this hypothesis, we computationally created artificial protein sequences using only the statistical information encoded in multiple sequence alignments extracted by the SCA, and built libraries of the designed sequences for experimental studies of folding and function. A study of many natural and artificial WW domain sequences shows that the statistical energy function capturing coupling between amino acid residues is necessary and sufficient to predict sequences that fold into native structures. The artificial proteins show thermodynamic stabilities and binding properties similar to natural WW domains, and structure determination of one artificial protein shows excellent agreement with the WW fold at atomic resolution. We suggest that the statistical energy function used for creating sequences represents evolution's rules for building a WW domain. The relative simplicity of the energy function suggests new testable hypotheses for understanding the physical chemistry mechanisms underlying protein folding and function.