Advances and Challenges in *De Novo* Protein Design

**Professor Christodoulous A. Floudas**
Department of Chemical Engineering
Princeton University

**Abstract**

The primary objective in *de novo* protein design is to determine the amino acid sequences which are compatible with existing or postulated template backbone structures that may be rigid or flexible. The *de novo* protein design problem is of fundamental importance since it addresses the mapping of the space of amino acid sequences to known protein folds or postulated/putative protein folds. It is also of significant practical importance since it can lead to the improved design of inhibitors, design of novel sequences with better stability, design of catalytic sites of enzymes, and drug discovery.

The first part of this lecture will provide a motivation for the *de novo* protein design problem, a definition of the flexible backbone template structures, and an overview of the advances and limitations. The second part will introduce a novel two-stage approach which takes into account explicitly the flexibility of the templates. The first stage addresses the *in silico sequence selection problem* through two key contributions: (a) the development of a distance-dependent Ca-Ca and side chain centroid-centroid distance dependent force fields; and (b) a rigorous quadratic assignment-like formulation for the prediction of a rank-ordered list of sequences with novel mutations. The second stage addresses the *fold specificity problem* by performing structure prediction calculations using atomistic level force fields. Two alternative approaches will be presented for the generation of ensembles of protein conformations: (i) the first principles protein structure prediction approach, Astro-Fold, and (ii) an approach motivated by an established NMR structure refinement protocol. Based on the ensembles of protein structures generated, the probabilities of each predicted sequence to fold specifically to the flexible templates are calculated. The theoretical prediction results for several peptides and proteins that include variants of Compstatin, human beta defensin-2, C3a, and gp41 for HIV-1 will be presented. Comparisons with experimental findings will also be discussed.