Protein interactions, unfolding and aggregation from low to high protein concentrations *via* coarse-grained molecular modeling and experimental characterization

Cesar O. Calero-Rubio
Advisor: Christopher J. Roberts
Committee Members: Eric M. Furst, Abraham M. Lenhoff, Atul Saluja

The diverse behavior of protein solutions can be attributed to the collection of microscopic interactions between solvent, protein, buffer and other cosolute molecules. These interactions can dictate different solution properties such as native and non-native aggregation, phase separation (liquid-liquid separation, crystallization, etc.), opalescence and elevated solution viscosity; and can be studied using a combination of different experimental and computational techniques. Despite current advances for computing infrastructure and the development of faster and more accurate experimental techniques, relatively little effort has been devoted to using molecular simulations to efficiently and accurately predict protein-protein interactions of concentrated protein solutions, as well as the thermodynamics of unfolding and aggregation, using coarse-grained (CG) molecular models. Furthermore, it is common to use molecular simulations in a "hindsight" manner, where the experimental behavior is already known and the simulations are intended to give molecular-scale insight that is beyond the capabilities of the experiment, or to help confirm or refute hypotheses that were based on interpretation of the experimental data. Much less work has been devoted to predicting experimental behavior with molecular models, either *a priori* or based on a subset of experimental data that can provide a reference for future predictions.

This presentation will focus on characterizing "weak" protein-protein, protein-cosolute and protein-solvent interactions, and protein unfolding behavior both experimentally and theoretically, as well as gaining insights in how formulation conditions mediate protein behavior in solution. Due to the significant differences in time and length scales of these molecular events, this presentation will be divided into three sections. The first section will focus on a methodology that combines experimental measurements at low protein concentrations (c_2) and molecular modeling to predict protein-protein interactions at higher c_2 values as a function of c_2, pH, and NaCl and other cosolutes concentrations. This is coupled with a revised framework to assess the effects of added cosolutes on protein solution behavior *via* density measurements and inverse Kirkwood-Buff solution theory. The second section will demonstrate how CG models can be used to better understand and/or predict the colloidal stability of protein solutions as a function of formulation conditions. Additional simulations of domain-domain preferential contacts will be explored for monoclonal antibodies. The last section will focus on the application of a refined implicit-solvent CG molecular model to predict the thermodynamics of unfolding and peptide-peptide self-association of a series of Ala-rich helical peptides as a function of peptide sequence, surface charge, hydrophobicity and chain length. The combination of all these approaches demonstrate the potential of coupling molecular scale computer simulations and robust analytical derivations with experimental training sets to accurately predict conformational and colloidal stability of protein and peptide solutions from low to high c_2. These new frameworks have the potential to be used as complementary tools during the discovery and development stages of biopharmaceuticals.