Large-scale characterization of binding landscapes using structural simulations: application to C_2H_2 zinc-finger transcription factors and MHC proteins

Chen Yanover and Philip Bradley
Program in Computational Biology,
Fred Hutchinson Cancer Research Center,
Seattle, WA
{cyanover,pbradley}@fhcrc.org

The prediction of macromolecular interactions is a key challenge for computational molecular biology. Given that molecular interactions are determined by the three-dimensional structures and chemical properties of the interacting partners, it seems plausible that such interactions could be predicted by structural modeling. One class of macromolecular interactions that represents a promising target for structure-based prediction consists of those interactions which are mediated by a linear sequence motif (peptide, DNA, RNA) in the partner molecule. For these interactions, the space of possible binding partners can be enumerated concisely, and approximate binding modes can often be inferred from the structures of related complexes. We describe a new approach for structure-based characterization of such motif-mediated interactions that integrates sampling techniques from protein-protein docking, loop modeling, and *de novo* structure prediction. The extensive conformational sampling is coupled with sequence space exploration to assemble, for each protein, a binding landscape consisting of sequences, molecular models, and predicted binding energies for an ensemble of optimized partners. We applied this protocol to large sets of (a) transcription factors in the C_2H_2 zinc-finger family and (b) class I major histocompatibility complex proteins, and demonstrated its efficacy in predicting protein-DNA and protein-peptide binding preferences. Importantly, analysis of the sequences, structures, and energies that comprise the landscapes provides new insights into structural determinants of binding specificity. Our simulation methodologies are implemented in the Rosetta molecular modeling package and can be generalized to other protein families.