Modeling of remote protein homologs via alignment ensembles with fragment-based perturbation

Dylan Chivian & David Baker

Dept. of Biochemistry, University of Washington, Seattle, WA 98102 USA
chivian@u.washington.edu, 206-543-7228

Whole genome sequencing projects have provided large numbers of gene sequences for which no structural information exists, creating a need for improved methods of protein structure prediction. When a structural homolog can be found, it is possible to build a model based on the coordinates of that homolog. Homology modeling typically presents four challenges: 1) detection of the closest structural homolog, 2) alignment of target residues to equivalent positions in the homolog, 3) modeling of loops and other structurally variable regions not provided by the homolog directly, and 4) refinement of models to make them more native-like. We present a method that combines these challenges in order to create model ensembles from which near-native models are selected by energy. Conformational variability is accomplished by the use of multiple homologs as templates and by parametric generation of alignment ensembles. Loops are added from fragments of experimental structures and modeled in the context of the template. Finally, in an effort to capture the sequence-induced local and global perturbations between the target and the template, fragment-insertion is performed along the entire length of the model, permitting escape from the frozen-approximation. Allowing many candidate conformations to follow the modeling process through to completion avoids early discards of superior solutions due to conformational flaws that may ultimately be remedied. We find that this approach performs well for obtaining near-native models on a set of difficult targets.