Representations of Structural Motifs for Protein Function Prediction

Brian Y. Chen[1], Drew H. Bryant[2], Joseph H. Bylund[2], Amanda E. Cruess[2], David M. Kristensen[3], Viacheslav Y. Fofanov[4], Mark Moll[1], Marek Kimmel[4], Olivier Lichtarge[4,5], Lydia E. Kavraki[1,2,5]

[1]Department of Computer Science, Rice University, [2]Department of Bioengineering, Rice University, [3]Program in Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine, [4]Department of Statistics, Rice University, [5]Department of Molecular and Human Genetics, Baylor College of Medicine

One strategy for function prediction is to search the structures of "target" proteins with unknown function for sites which are geometrically and chemically similar to "motifs" representing a known active site. Like all function prediction strategies, motif matching can have some inaccuracies, such as in the design of the motifs, which may have geometric and chemical dissimilarities to functionally related proteins (insensitive), or similarities to functionally unrelated proteins (unspecific). This poster describes two effective motif representations.

First, we describe composite motifs, which combine protein geometry from multiple functionally related active sites. Combining multiple structures allows composite motifs to capture active site variations not apparent in conventional single structure motif designs. In leave-one-out experimentation, composite motifs exhibit sensitivity among the highest in a group of single-structure motifs, while maintaining average specificity.

We then describe cavity-aware motifs, which integrate atom geometry with C–spheres that represent cleft and cavity volumes essential for protein function. We use C–spheres as an exclusion filter. In testing potential matches, we insist that target sites matching the atomic geometry of the motif also contain empty volumes matching the position of the C–spheres. This eliminates targets which do not have clefts and cavities similar to known, functionally significant volumes. In our experimentation, we demonstrate that cavity-aware motifs eliminate over 80% of matches to functionally unrelated proteins, while accidentally eliminating less than 20% of matches to related proteins.

Composite motifs and cavity-aware motifs provide two orthogonal techniques for improving motif sensitivity and specificity within the broader strategy of motif matching for protein function prediction.