

High Confidence Structure Prediction Using M-ran (Modification Reactivity Analysis)

G.M. Alter and M. Raymer

The structure of a protein provides it with its unique biological properties. Therefore a reliable understanding of that structure is a key step to understanding its various functions and intervening in those functions. A tremendous effort has been devoted to efficiently determining protein structures. Part of this effort is prediction of structure based on amino acid sequence. Formulating accurate predictions has proved to be a formidable task. As part of the effort to accurately predict structures we have combined structure prediction methodology with observed reactivity toward various chemical and enzymatic reagents, at multiple sites on protein structures. (We term this M-ran, or Modification Reactivity Analysis.) This approach depends on reasonable structure predictions. Generally our strategies to achieve these predictions involve identifying domains and using homology methods to model them or using ab initio methods for modeling regions that are not domains. For proteins with multiple regions, we have explored strategies for combining the regions (just mentioned) using a fast docking algorithm. We have investigated the use of energy-based methods including energy minimization, molecular dynamic simulations, and simulated annealing to refine predicted structures. Our strategy calls for evaluating adequacy of results using M-ran at multiple stages during model construction.

Reaction rates at multiple sites in a protein structure are extracted from time courses derived from mass spectral data that characterize the process of protein modification. To efficiently accomplish this we have developed software that assigns MS peaks, correlates and scales these peaks among multiple spectra, assigns (aids in assigning) the sites of these reactions, and generates time courses. We have focused on methods for analyzing time courses in terms of rate constants. These rate constants are the fundamental experimental measurement of M-ran. We take them as a measure of the reactivity at each site in a protein structure. (This property reflects the chemistry of specific reactions, and the detailed environment of the reactive site.) Guided by reactivities determined in known structures, we interpret reactivities in terms of the environment of each reactive group of a particular protein examined.

We have examined the structure of native Replication Protein A (RPA) under physiologically relevant conditions using the strategies just described. RPA is a heterotrimeric protein consisting of one 70, 32, and 14 kD subunit. It is an essential participant in DNA replication, recombination, and repair. We have distinguished among alternative reasonable structure models for RPA domains and inter-domain regions, and have selected those most relevant to the native protein, as judged by M-ran. Combining M-ran and energetic (docking) studies, we have constructed a model (family of models) for the complete protein. In this process we have addressed methods for handling coil, and perhaps, disorganized portions of protein structure. Our results provide a model that gives insight into the biological function of RPA. Further, they demonstrate the feasibility of using structure prediction combined with M-ran for constructing high confidence predictions of protein structures.