THE RELATIONSHIP BETWEEN PROTEIN STRUCTURE AND FUNCTION, OR HOW HAVE PROTEINS OVER TIME DIVERGED IN FUNCTION?

PATRICIA C. BABBITT
Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry
University of California, San Francisco, CA 94611

Protein structure-function relationships can be investigated by asking how nature has re-engineered protein structures to perform a variety of functions. Computational methods directed at the identification and analysis of related protein structures are an important prerequisite in this endeavor. The papers presented in this seminar track offer some new ideas for addressing fundamental problems in describing the structural elements of the structure/function paradigm. The many-faceted and disparate nature of the problem is sampled by the session which offers several approaches to understanding evolutionary conservation/ variation from sequence and structural information.

Understanding the relationships between protein structure and function remains a primary focus in structural biology with important consequences in such diverse fields as molecular biology, genetics, biochemistry, protein engineering and bioinformatics. One approach to this problem is to study how nature has re-engineered proteins for new functions through evolutionary processes. This strategy has the potential to reveal fundamental characteristics of protein structures and the explicit manner in which they deliver their associated functions.

To understand the structure-function paradigm, particularly useful structural information comes from the primary amino acid sequences and the associated tertiary structures. Several recent developments in analysis of the "protein universe" at the tertiary structural level have provided important criteria for understanding the range of family folds that exist and some of the evolutionary relationships associated with them. While the tertiary structure database is small, the sequence databases are large and now include the sequences of the entire genomes of several bacteria, an archaeon and a microbial eukaryote. Many additional genomes will be solved in the near future. Using the great deal of protein sequence and structural data at hand, computational strategies to address the "structure-function" problem can now support serious attempts to understand the fundamental relationships between protein structure and function.

The papers in this section represent a cross-section of the broad variety of issues that are pertinent to the area. While the dependence of the statistical significance of an alignment on the length of the aligned sequences has been addressed by many investigators, Alexandrov and Solovyev present a preliminary analysis that suggests that statistical significance of a local sequence alignment depends on the length of

the alignment itself. This observation could be used to enhance detection of distantly related proteins from database searches. The work presented by Dunker *et al*. addresses a much different issue: the problem of how protein structures evolve to confer specificity and affinity. This work flags disordered regions in crystal structures as data that can be evaluated to gain insight into that problem. On a more phenomenological level, Fetrow and Godzik present a hypothesis for how modern-day proteins may have evolved from much smaller "proto-proteins." Re-visiting an old suggestion that present-day proteins evolved from small peptides that exhibit secondary structure and primitive function, these authors re-interpret such observations to suggest that local structural similarities in some presumably unrelated proteins actually reflect divergent rather than convergent evolution. The papers by Wei and Altman and Wu *et al*. describe new approaches to structural analysis likely to be useful for understanding how protein architectures evolve to accommodate a variety of functions. The former paper describes an algorithm to produce a statistical description of spatial properties, such as metal binding sites, in protein structures. The output includes a score reflecting the likelihood that the region identified is indeed a site of interest. This approach could be developed for recognizing functionally relevant structures in unannotated sequences and for flagging very distant structural similarities. The paper by Wu *et al*. describes an approach for analyzing multiple related structures using an affine model and transformation matrices. The paper also introduces a new method for finding structurally corresponding regions by matching curvatures along the protein backbones. These approaches have potential for distinguishing relative conservation and divergence in local regions of related structures, an important prerequisite for understanding how proteins evolve to confer specificity or entirely new functions.

Acknowledgments

Thanks to those who contributed to this session by submitting their work or reviewing submissions.