THE EVOLUTION OF BIOMOLECULAR STRUCTURES AND THE STRUCTURE OF BIOMOLECULAR EVOLUTION

R. A. GOLDSTEIN
Biophysics Research Division
Department of Chemistry
University of Michigan
Ann Arbor, MI 48109-1055

R. B. ALTMAN
Section on Medical Informatics
Stanford University Medical Center
Stanford, CA 94305-5479

All biological macromolecules have an evolutionary origin that is manifested in essentially every aspect of their biological and physico-chemical properties. In order to understand these macromolecules, it is necessary to understand the relationship between these properties and the evolutionary pressures that determined form and function. Information about the evolutionary history can be used to help predict the characteristics of these macromolecules. Conversely, these macromolecules encode this evolutionary heritage, and can provide insight into the process of molecular evolution. This conference track is directed towards using computational methods to explore the relationship between such macromolecules and their evolutionary context.

At the initial stages of the evolution, the relatively simple precursors to modern cells must have performed all of the functions necessary for survival. In the absence of laboratory specimens of early life, it is difficult to study how such "proto-cells" could have fulfilled these tasks. Andrew Pohorille and co-workers describe their work using large-scale molecular dynamics to model the circumstances under which various membrane-associated processes might have occurred, and what molecular machinery would be required. Such work can help to establish both necessary and sufficient properties of these early organisms.

In his paper, Erich Bornberg-Bauer describes his work using a different type of computer modeling. He constructs large ensembles of RNA and protein sequences, and by using simple models to simulate how these sequences would fold, investigates the relationship between biopolymer sequence and structure. These studies provide insight into the fitness landscapes underlying molecular evolution, and helps to answer how evolutionary-constrained biopolymers can be both functionally adaptable and structurally robust.

The increasing abundance of sequence data allows us to look at sequential meta-phenomena, such as trends in the overall sizes and amino acid distributions of protein molecules, providing important clues to evolutionary heritage and constraints on structure/function relationships. In their paper, Kolker and Trifinov study periodicities in protein sequence length and find evidence for protein subunit "quanta" of sizes 123 amino acids (for eukaryotes) and 152 amino acids (for prokaryotes). They discuss the implications of this observation for understanding the evolutionary pressures on protein structure.

Understanding how to relate the observed rates of site mutations in biological proteins to the corresponding changes in the physico-chemical properties of the amino acids is complicated by the fact that the protein must fulfill a large number of criteria, including stability, functionality, and foldability. In their contribution, Jeffrey Koshi and Richard Goldstein use their recently-developed structure-dependent optimal substitution matrices to try to understand what properties are conserved during the process of evolution, how this depends on local structure, and what this indicates about the necessary conditions for biological activity.

Most research into evolutionary relationships starts with the construction of a phylogenetic tree that represents the evolutionary relationships between biological units. The construction of phylogenetic trees representing the evolution of proteins based on their animo acid sequences remains a difficult problem due to the number of possible trees for even a moderate number of homologous proteins, and the possible presence of significantly different, nearly-optimal solutions. In his paper, Hideo Matsuda describes how he employs genetic algorithms to search for optimal phylogenetic trees.

Obtaining accurate data regarding the number and types of protein families, both within and between species, is a critical step in understanding the evolutionary relationship structure and function. Methods are needed to classify sequences into families as rapidly as the sequences are produced. Wu and coworkers report a method, based on neural networks, which is able to identify sequence motifs with promising specificity and sensitivity, and at rates that are substantially faster than some of the more commonly used sequence comparison tools.

Acknowledgments

We would like to thank all of the researchers who submitted work to this track, and for all of the reviewers who found time to provide many helpful comments and criticisms.