

Computational and analytics challenges in biology

Silvia Crivelli, University of California, Davis

One of the biggest challenges in modern biology is to develop an understanding of biological processes to the level that it would turn biology into a predictive science. Understanding these complex biological systems at different levels, from cellular to molecular and all the way to tissue and organ level, requires multidisciplinary teams that bring together applied mathematicians, highperformance computing scientists, and domain scientists. In this talk, I will describe two cases that clearly highlight this challenge. The first case is the protein structure prediction problem, which has eluded scientists for more than 50 years. Protein structure prediction is to efficiently determine a protein's 3D structure – its native structure -- from its primary sequence of amino acids using computational methods. Challenges in protein folding and refinement include global optimization of rugged functions with many degrees of freedom, efficient sampling of a protein conformational space, and ranking of hundreds of thousands of models that may have subtle variations. The second case is the simplified geometric representation of large and complex 3D images of biological systems such as the cytoskeleton. These images are too large and complex to interpret and visualize in their entirety forcing biologists to segment them into sub-volumes thus missing important information and losing global perspective. Geometric representations will allow biologists to create models of those systems which can be used to see common local and global patterns, integrate biological information from different modalities, and run simulations which can help them to quickly conduct virtual lab tests and formulate hypothesis that they can then try on the wet lab. Challenges in bio-imaging include rendering and interpreting very large and complex data sets, achieving a smooth surface simplification using less storage, and training computers to “learn” how to recognize similar cartoons once they have been identified by a human.