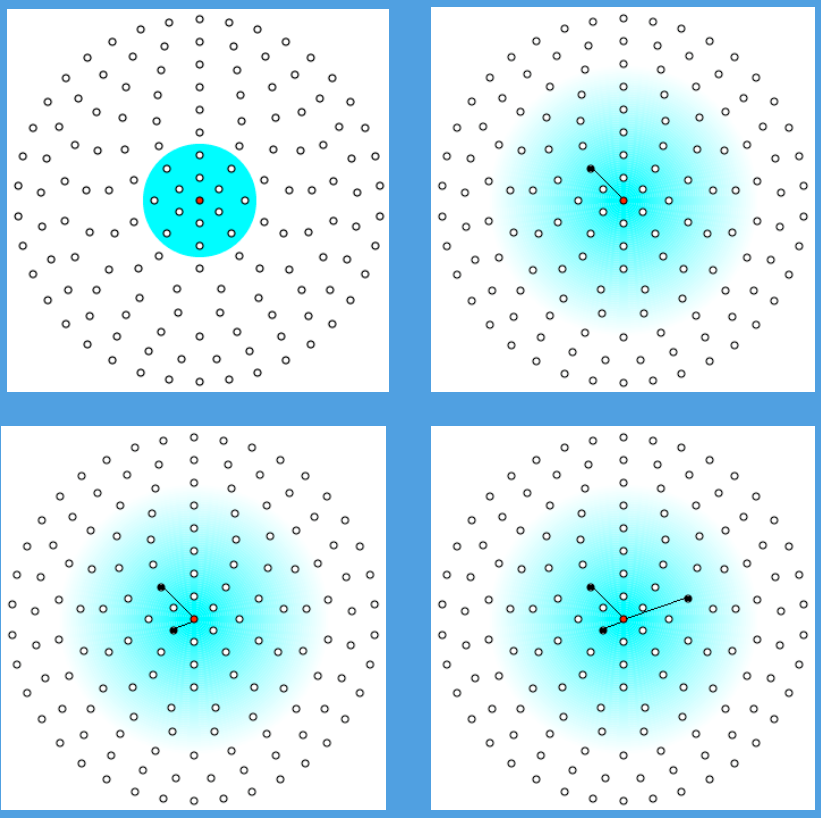
Brian Cloteaux
Nadezha Serova
(UMBC)

complex systems

IMAGE OF THE MONTH

April

Matching Observed Alpha Helix Lengths to Predicted Secondary Structure



Algorithm randomly choosing possible helix orders in a permutation space where the blue denotes a selection

More information available at: http://www.itl.nist.gov/ITLPrograms/ComplexSystems/

Matching Observed Lengths to Predicted Structure

We are interested in generating a library of probable matchings of observed alpha helix lengths to 1D structure.
The idea is to provide a first attempt at understanding the 3D structure.

Measuring distance between orders

We used Kendall-tau distance to compare orders. As shown our heuristic method often produces better results to the actual order than an optimal order.

Extending a base order to a library

Starting from a base order π, we create a random permutation σ with a probability proportional $(1\text{-}p)^{d_{K\tau}(\pi,\sigma)}$ for a parameter p.

Testing the method

We generated 200 random permutations and measured distance to actual protein order. The results show that every time either the exact or a close order was produced.

Because of the complexity in determining the 3D structure of a protein, the use of partial information determined from experimental techniques can greatly reduce the overall computational expense. We examined the problem of matching observed lengths of alpha helices to their predicted location on a protein's amino acid sequence. This potentially can be a first step towards determining the 3D structure of the protein.

We showed that the effort in finding optimal potential solutions does not seem to be worth the computational expense. In particular, we showed evidence that, because of the uncertainty of the helix prediction, the optimal coverings can be relatively distance from the actual ordering on the protein. Instead, we introduced a simple greedy heuristic for estimating the order. Using this heuristic as a starting point, we chose random orders around it using the BubbleSearch method of Lesh and Mitzenmacher. When compared to the actual orderings, this method was able to either find the correct ordering or an ordering that was very close. Thus, we believe that our method is a fast and efficient algorithm for determining a set of potential placements of helix lengths onto a protein sequence.

complex systems
Information Technology Laboratory

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

2010

The Complex Systems Program is part of the National Institute of Standards and Technology's Information Technology Laboratory. Complex Systems are composed of large interrelated, interacting entities which taken together, exhibit macroscopic behavior which is not predictable by examination of the individual entities. The Complex Systems program seeks to understand the fundamental science of these systems and develop rigorous descriptions (analytic, statistical, or semantic) that enable prediction and control of their behavior.

Program information at: `www.itl.nist.gov/ITLPrograms/ComplexSystems`