a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA

Graduate Program in
BIOINFORMATICS

100-570 West 7th Avenue
Vancouver, B.C. Canada V5Z 4S6
Tel: (604) 707-5803
Fax: (604) 876-3561
http://bioinformatics.ubc.ca

M.Sc. EXIT SEMINAR

Jeff Proctor

BSEng, University of Victoria, 2010

Friday, September 21st, 2012 at 11 AM

LOCATION: Lecture Theatre, BCCRC

CoFold: an RNA structure prediction method that takes co-transcriptional folding into account

Abstract:
RNA has a diverse array of cellular functions, and relies on molecular structure to carry them out. The vast majority of current methods for prediction of RNA secondary structure (i.e. the set of base pairs in the molecule) consider the minimum free energy structure (i.e. the most thermodynamically stable structure), and thus disregard the process of structural formation. There exists substantial evidence that the process of structure formation is important, and that it does impact the resulting functional RNA structure. Several methods currently exist that explicitly simulate the kinetic folding pathway as a time-ordered sequence of structural changes. However, these methods not only suffer from a long list of limiting assumptions about the cellular environment, but also are restricted to short sequences.

In this work, we explore the idea of capturing the effects of kinetic folding rather than simulating in detail the process over time, and propose that accounting for kinetic effects of structural formation is crucial to further improve non-comparative RNA secondary structure prediction methods. During transcription, RNA structure begins to form immediately as the molecule emerges from the polymerase (i.e. co-transcriptionally). Long-range base pairs suffer a disadvantage during this process, as quickly forming short-range base pairs act to block their formation (i.e. due to kinetic barriers).

We propose a novel method, CoFold, that captures the reachability of potential pairing partners during co-transcriptional folding. We show that it significantly improves prediction accuracy over free energy minimization alone, particularly for long sequences. CoFold depends on only two free parameters that are highly correlated, and we demonstrate robust training. Furthermore, the resulting structures predicted by CoFold have a free energy measurement that does not significantly differ from that of the respective RNAfold prediction, indicating that they are indeed stable secondary structures. We propose that consideration of kinetics in RNA secondary structure prediction is crucial, and we hope that this work encourages further exploration of its effect on biological RNA structure.

Supervisor: Dr. Irmtraud Meyer, Assistant Professor, Department of Computer Science, UBC