

An investigation of structural profiles around target sites of RNA binding proteins.

Tsukasa Fukunaga¹, Hisanori Kiryu¹,

¹ The University of Tokyo, Japan

Keyword: RNA binding proteins, RNA secondary structure, CLIP-seq

Background and Motivations

RNA binding proteins (RBPs) play an integral role in post-transcriptional regulation by binding to target transcripts. Therefore, an understanding of RBPs and their target sites lead us to understanding gene regulatory networks. Recent improvement of experimental technologies including CLIP-seq enables us to identify target sites of RBPs comprehensively. Thus, we can discover the sequential motif of the CLIP-ed RBP. However, RBPs recognize not only specific sequential motifs but also RNA secondary structure in their target sites. Despite the importance of the RNA secondary structure in target sites of RBPs, there have been few studies investigating RNA secondary structures in target sites. When an RNA sequence forms a secondary structure, each nucleotide in the sequence takes any of six different loops, bulge loop, hairpin loop, internal loop, multi loop, outer loop and stem. Here, we define that the structural profile of nucleotide i in an RNA sequence are probabilities that i takes each six loop type. Although we need structural profiles in order to detect binding to specific loop type, none of the existing programs can exactly compute the structural profiles.

Proposed Approaches

Therefore, we developed algorithms for computing the structural profiles exactly using a dynamic programming method, and implemented in software called 'CapR'. We computed the structural profiles based on the Rfold model. The Rfold model is an unambiguous grammar that generates all the secondary structures excluding pseudoknots without redundancy. This model is also able to be directly applied to the energy model and we apply Turner Energy model. To apply the algorithm to long sequences, we restricted the maximal span of the base pairs to a fixed value W . The computational complexities of our algorithm are $O(NW^2)$.

Results and Conclusions

First, we analyzed the basic properties of structural profiles. We found that maximal span and GC content have an impact on structural profiles. Second, we evaluated accuracy of structural profiles calculated by CapR. We showed that our method accurately infers loop types. Third, we investigated structural profiles of target sites of RBPs determined by RIP-Chip and CLIP-seq. We verified that yeast protein Vts1p preferentially bind hairpin loops in vivo. Also, structural profiles are more useful than accessibilities in order to distinguish the false binding site from the true binding site. Last, we investigate the structurally most important positions around binding sites of RBPs. While neighboring binding site is important for the QKI protein, 5'-end of their binding site is important for the Nova protein. These results may indicate the kinetic aspects of the binding mechanism.