STONE: an Application to Investigate Protein Secondary Structure Transitions

Armano G[(1)], Manconi A[(1)]

[(1)] Department of Electrical and Electronic Engineering, University of Cagliari, Cagliari

Motivation

Predicting protein secondary structures is an important task in bioinformatics, and different approaches have been proposed to deal with it. To simplify, we can consider last generation predictors as a pipeline of three functional abstractions:
i) encoding,
ii) primary-to-secondary structure prediction, and
iii) post-processing.
We believe that interesting margins of improvement can be found in the post-processing phase.

Methods

Assessing the behavior of existing predictors can be useful to find out tips for implementing effective post-processing modules. In particular, since a significant part of the prediction errors are associated with transitions, we deem that algorithms aimed at correcting secondary structure transitions might improve the performance of a predictor to some extent. The most acknowledged way of representing secondary structure uses three symbols: h, e, and c to denote amino acids lying within alpha helices, beta-strands, or coils. Nevertheless, if one wants to study transitions, other representations appear more suitable. In particular, we adhered to on-off diagrams to represent alpha-helices and beta-strands (coils are implicitly defined). To better understand the ability of a predictor in the task of predicting secondary structure transitions, we redefined also precision and recall. In particular, we introduced an extra parameter, i.e. granularity (G), which allows to disregard errors when predicted transitions are sufficiently close to the actual ones. For instance, with G = ±2, a prediction will be considered erroneous only when the actual transition occurs outside a window of 5 amino acids (centered on the predicted transition). Let us point out that predicted transitions are typically close to the actual ones, allowing to formulate the conjecture that local post-processing performed on predicted transitions can be effective.

Results

To help researchers better understand the ability of a predictor to deal with secondary structure transitions, we developed STONE (Structural TransitiONs Evaluator). Implemented in Python, STONE is a stand-alone application that currently allows:
i) to display actual and predicted transitions of alpha-helices and beta-strands, given a predictor and a target protein,
ii) to compute and display granular precision and recall, given a predictor and a set of proteins, and
iii) to compare the performance of different predictors on the same set of proteins in terms of granular precision and recall.
Preliminary experiments have been performed running six secondary structure predictors: GOR IV, Prof, Predator, SSPRO, JNET, and PSI-PRED on relevant test sets (including RS126). Results show that only 35% of transitions, on average, are properly predicted. Taking as a reference the Q3 index evaluated with G=0, we found upper bounds for the expected improvement obtainable by running a post-processor that acts locally on transitions. In particular, our experimental results show that –on average– repairing all errors within a window of 3 amino acids (G=±1) would give an improvement of 3%, whereas an improvement of 5% is expected with a window of 5 amino acids (G= ±2). As for future work, we are planning to investigate the bias in predicting transitions, i.e. the propensity of a predictor to anticipate or postpone them. We deem that having information about the presence of a bias will allow to write simple and effective procedures able to improve the Q3 index.

Contact : armano@diee.unica.it