ABSTRACT

HARYANTO, TOTO. The Development of Hidden Semi Markov Model with Empirical State Duration Distribution for Protein Secondary Structure Predictions. Under direction of AGUS BUONO and ANTO SATRIYO NUGROHO

This research aimed to develop Hidden Semi Markov Model (HSMM) with long-duration distribution of state empirically and theoretically to predict protein secondary structure. Data used in this study are subset data taken from database of secondary protein structure in DSSP program with three secondary protein structures of alpha-helix (H), beta-sheet (B), and coil (C). Accuracy of the predicted protein structure with HSMM is compared with Hidden Markov Model (HMM) standard. The results showed that the HSMM generally provides a prediction accuracy 72.1%, 35.3% and 63.1% for H, B and C respectively. The accuracy of standard HMM is 98.4% for (B), 1% for (B) and 6.4% for (C). The use of empirical state duration distribution gives better accuracy on average compared a theoretical state duration distribution. Use of the empirical state duration distribution provides accuracy 71.2% for (H), 37.1% for (B) and 64.0% for (C). Use of theoretical state duration distribution provides accuracy 75% for (H), 30.2% for (B) and 60.6% for (C). Less accuracy in beta-sheet prediction with HSMM was caused by high of the dispute between distribution in training data and testing data.

Keyword: Hidden Semi Markov Model (HSMM), Hidden Markov Model (HMM), protein secondary structure