P19

Randomized Subspace Learning for Proline Cis-Trans Isomerization Prediction

O. Y. Al-Jarrah 1, P. D. Yoo 1, K. Taha 1, S. Muhaidat 1, A. Shami 2, N. Zaki 3

1 ECE Dept., Khalifa University, Abu Dhabi

2 ECE Dept., University of Western Ontario, Canada

3 College of Information Technology, UAE University, UAE

Abstract

Proline residues are common source of kinetic complications during folding. The X-Pro peptide bond is the only peptide bond for which the stability of the cis and trans conformations is comparable. The cis–trans isomerization (CTI) of X-Pro peptide bonds is a widely recognized rate-limiting factor, which can not only induces additional slow phases in protein folding but also modifies the millisecond and sub-millisecond dynamics of the protein. An accurate computational prediction of proline CTI is of great importance for the understanding of protein folding, splicing, cell signaling, and transmembrane active transport in both the human body and animals. In our earlier work, we successfully developed a biophysically motivated proline CTI predictor utilizing a novel tree-based consensus model with a powerful metalearning technique and achieved 86.58% Q2 accuracy and 0.74 Mcc, which is a better result than the results (70–73% Q2 accuracies) reported in the literature on the wellreferenced benchmark dataset. In this paper, we describe experiments with novel randomized subspace learning and bootstrap seeding techniques as an extension to our earlier work, the consensus models as well as entropy-based learning methods, to obtain better accuracy through a precise and robust learning scheme for proline CTI prediction.