



ID de Contribution: 101

Type: Non spécifié

## An SVM-Domain Linker Prediction Trained with Optimized Features Selected by Random Forest and Stepwise Selection

Computer aided/assisted methods for predicting structural domains and their boundaries are actively investigated because biologically significant proteins are often large, multi-domain proteins and difficult to be characterized by high-throughput methods. Thus, efficient methods for predicting domain/boundary are becoming of practical importance in diverse area of proteomics project.

Here, we constructed a new support vector machine (SVM) based domain linker predictor, DROP (Domain linker pRediction using OPTimal features). The SVM was trained with 43 optimized features. The optimal feature combination was selected from 2870 features by using a random forest algorithm complemented with a backward feature selection. The computation time of the random forest and the backward selection were, respectively, 20 hours and 100 hours on an 8 Xeon processors Linux server. The prediction sensitivity and specificity of DROP were 49.5% and 62.6%, respectively. These values were higher than those of control SVM predictors trained with other feature combinations, strongly suggesting that our feature selection method indeed selected optimal features. In addition, DROP demonstrated the highest NDO-Score among the 12 CASP8 DP servers when assessed with 7 CASP8 FM multi-domain targets.

These results indicate that the SVM prediction of domain linkers can be improved by identifying the optimal combinations of features that best distinguish linker from non-linker regions. DROP, constructed on such premise, can efficiently predict domain linker regions in novel multi-domain protein targets, even when the linker sequence signatures are weak.

**Auteur principal:** M. EBINA, Teppei (Dept of Biotech and Life Sci, Tokyo Univ of A & T (TUAT))

**Co-auteurs:** Dr TOH, Hiroyuki (Div. Bioinf, Med. Inst. of Bioreg, Kyushu Univ); Dr KURODA, Yutaka (Dept of Biotech and Life Sci, Tokyo Univ of A & T (TUAT))

**Orateur:** M. EBINA, Teppei (Dept of Biotech and Life Sci, Tokyo Univ of A & T (TUAT))