１．　課題名

Adaptation of local clustering techniques to biomedical data sets

２．　代表者名

HOULE Michael（国立情報学研究所）

３．　研究成果の概要

○　Development of a data preparation method for protein sequences and an associated similarity measure supporting accurate and efficient similarity search (using the SASH search structure due to Houle & Sakuma, 2005).

○　Development of an enhanced local clustering model, and implementation of a prototype local clustering tool for protein sequence data based on this model, using the aforementioned data preparation method (completed June 2005).

○　Testing of the prototype clustering tool on a set of 378,659 bacterial protein sequences. The clustering was successful in a technical sense: 19100 clusters ranging in size from 4 sequences to over 14000 sequences were produced in 13 hours of computation on a desktop computer, after data preparation of 1 day on a 16-node PC cluster. The semantic quality of the clustering is still under evaluation.

○　We expect to submit papers on these research results in H17.