PSC: An human-computer interaction system for Protein 3D Structure Comparison

X. Chen, Y. Q. Chen, S. M. He, W. Gao
Institute of Computing Technology, Chinese Academy of Sciences, P.R.China,100080
Contact: E-mail: xchen@jdl.ac.cn

1. Motivation:

Since protein 3D structures are more conserved than sequences and even functions, analyzing the protein 3D structure is very important to the understanding of biological processes. With the accumulation of Protein structure information, attention has been paid to the development of methods for comparing and classifying those structures.

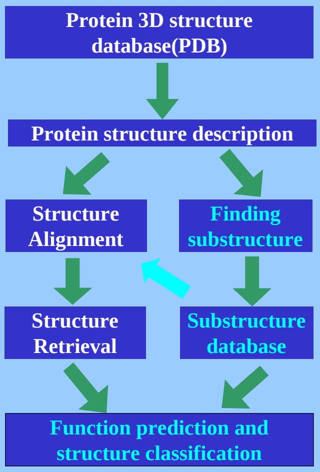
2. Method Introduction:

A system of comparing and retrieval protein 3D structures is introduced here. The system applies the moment feature extraction technology and the feature subspace technology in pattern recognition to identify and compare different protein structures.

And the system can be optimized by some of the current protein structure classification database. Furthermore, the system offers Effective Human-Computer Interaction that can dynamically improve the system through the biologist's feedback on interface in real-time, which is attracting to biologist. Web-based Dali server, which performs well in 3D structure matching, is used as the ground truth to make assessment of our system.

In our experiments on protein structure datasets, our system performs close to the ground truth with much simplicity and efficiency. And through biologist's feedback on interface, the system can improve over 10% accuracy on average.

3. Some charts illustrate the system:


Protein 3D structure database(PDB)

Protein structure description

Structure Alignment

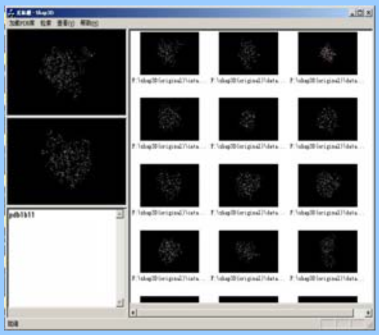
Finding substructure

Structure Retrieval

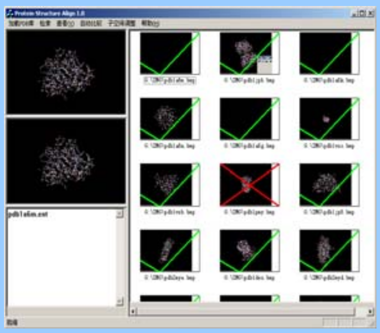
Substructure database

Function prediction and structure classification

1. *Prospect of our system*



2. *The interface of our system(after retrieval)*



3. *The interface of our system(after relevant feedback)*

Representative Protein ID	Initial Matching (Retrieval Accuracy)	First Relevant Feedback (Retrieval Accuracy)	Second Relevant Feedback (Retrieval Accuracy)
1cxq	80%	90%	100%
1a6m	80%	90%	95%
1abw	90%	100%	
1ba1	85%	100%	
1b8j	70%	75%	75%
1ctq	45%	65%	75%
1djx	40%	45%	55%
1fnc	35%	40%	40%
1crb	20%	25%	30%
1ckq	30%	35%	40%

4. *Retrieval Accuracy of different representative sets*

ACKNOWLEDGEMENTS

This work was supported by the National Key Basic Re-search & Development Program (973) of China under Grant No. 2002CB713807.