

# MPSS: an integrated database system for surveying multitudinous proteins

**Pei Hao**<sup>1,2</sup>  
phao@sibs.ac.cn

**Wei-zhong He**<sup>2</sup>  
hewz@sibit.org

**Yi-xue Li**<sup>1,2</sup>  
yxli@sibs.ac.cn

<sup>1</sup> Bioinformation Center, Shanghai Institute for Biological Sciences, CAS, 320# Yueyang Road, Shanghai 200031, China

<sup>2</sup> Shanghai Center for Bioinformation Technology, 100, Qinzhou Rd, Fl.12, Shanghai 200235, China

**Keywords:** MPSS, database, highly instructive

## 1 Introduction

The **M**ultitudinous **P**rotein **S**urvey **S**ystem (MPSS) provides an integrated, user-friendly platform to retrieve information of multiple proteins at one time. This system integrates several important and widely used databases including SwissProt, TrEMBL, GO, KEGG and contains many useful references to other databases such as PDB, Pfam etc. Since most of online available protein information retrieve databases can only process one protein at a time, it is a laborious and time-consuming for researchers to submit their large number of queries one by one and wait for the results. MPSS can greatly simplify this kind of retrieve procedure. Researchers can submit a group of protein IDs or entry names sourcing of SwissProt and TrEMBL to MPSS and get the annotated protein information directly. Therefore, retrieving protein information from MPSS can reduce the query time significantly. In addition, MPSS provides users comprehensive information of target proteins, which includes 3D structures, domains, protein-protein interaction, pathway and function knowledge based on gene ontology. The information highly instructive for researchers to have a comprehensive view of protein functions they retrieved.

## 2 Method and Results

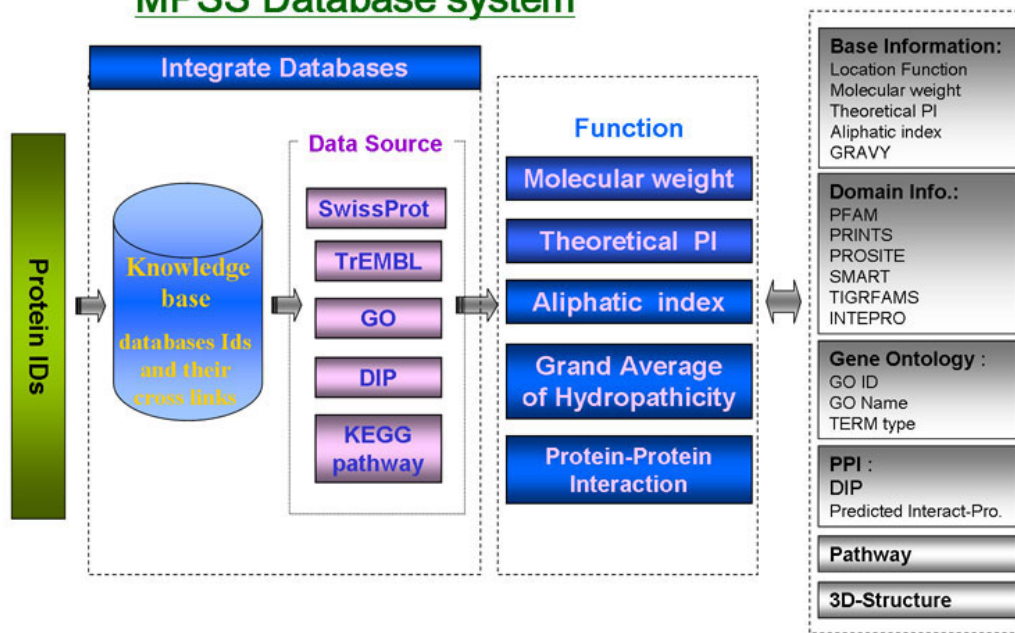
### 2.2 Online Accessibility

MPSS is available as a Web-based service at <http://www.biosino.org/MPSS/index.jsp>.

### 2.2 Figures

**Figure1.** The overview of the flow chart of MPSS. The normalized protein IDs are mapped to related information in MPSS. Data sources include SwissProt, TrEMBL, GO, DIP, KEGG. The main functions of MPSS are listed on the right side of the figure.

## MPSS Database system



### 3 Discussions

Given the rapid advance in biological sciences, the ability to integrate new data and adapt to new data format is necessary to the success of the design of the information service system. The simple elegant internal database structure of MPSS makes it an easy task to include other important protein information, if proved to be very useful to users in the future. As a result, researchers will be able to retrieve more and more valuable information from MPSS.

Researchers in proteomics and microarray, who wants to spend more time on other interesting and valuable biological questions, have already experienced the benefits of this flexible batch-working approach of MPSS in their daily work. Future work will be focused on how to organize the data in MPSS more efficiently and customize the services to fit different users.

We believe that this new concept approach of acquiring molecular information will become a trend in querying database considering the increasing data amount and growing complexity involved in bench work, thus, our ultimate goal is to turn MPSS into a fully automatic pipeline for researchers to retrieve protein information. To provide more preprocessed information such as the reconstructed local protein network, the predicted function, interaction or even 3D structure is also in our consideration.

### References

1. Magrane M., Apweiler R. (2002) Organization and Standardization of information in Swiss-Prot and TrEMBL. *Data Science Journal*, **1**(1), 13-18.
2. Boeckmann B., Bairoch A., Apweiler R., Blatter M., Estreicher A., Gasteiger E., Martin M. J., Michoud K., O'Donovan C., Phan I., Pilbout S., and Schneider M. (2003) The Swiss-Prot protein knowledgebase and its supplement TrEMBL. (2003). *Nucl. Acids. Res.*, **31**, 365-370 .