SubLoc: a server/client suite for protein subcellular location based on SOAP

Hu Chen, Ni Huang and Zhirong Sun*

Institute of Bioinformatics and Systems Biology, MOE Key Laboratory of Bioinformatics, State Key Laboratory of Biomembrane and Membrane Biotechnology, Department of Biological Sciences and Biotechnology, Tsinghua University, Beijing 100084, China

Received on September 9, 2005; revised on November 11, 2005; accepted on December 5, 2005

Advance Access publication December 8, 2005

Associate Editor: Alvis Brazma

Databases and ontologies

SubLoc: a server/client suite for protein subcellular location based on SOAP

Hu Chen, Ni Huang and Zhirong Sun*

Institute of Bioinformatics and Systems Biology, MOE Key Laboratory of Bioinformatics, State Key Laboratory of Biomembrane and Membrane Biotechnology, Department of Biological Sciences and Biotechnology, Tsinghua University, Beijing 100084, China

Received on September 9, 2005; revised on November 11, 2005; accepted on December 5, 2005

Advance Access publication December 8, 2005

Associate Editor: Alvis Brazma

ABSTRACT

Summary: Based on SOAP (Simple Object Access Protocol) technology, the SubLoc server/client suite offers a user-friendly interface for searching and predicting protein subcellular location.


Contact: sunzhr@mail.tsinghua.edu.cn

Supplementary information: Supplementary data are available at Bioinformatics online.

INTRODUCTION

Proteins are sorted into different cellular compartments such as cytoplasm, nuclear region, mitochondrion, etc. or may be secreted out of the cell, and their proper functioning relies on this precise process of subcellular localization. Hence, subcellular location information may imply the function. DBSubLoc (Guo et al., 2004) is a protein subcellular localization annotation database, which is available at http://www.bioinfo.tsinghua.edu.cn/dsubloc.html. The database contains >60 000 protein sequences from virus, bacteria, fungi, plant and animal. However, its service is via WWW and the user interface is built on web browsers, which is designed to be accessed by humans, not by machines. Thus, it is troublesome for users to use DBSubLoc in an automated manner.

We adapted the SOAP (Simple Object Access Protocol. http://www.w3.org/TR/soap) technology to DBSubLoc database and developed a server/client suite to facilitate automated information retrieval. SOAP uses XML to define an extensible messaging framework, providing a message construct that can be exchanged over a variety of underlying protocols. Because of its simplicity and extensibility, it is being widely applied to more and more bioinformatics databases such as KEGG (Kanehisa et al., 2004; http://www.genome.jp/kegg-soap) and NCBI Entrez Utilities Web Service (Maglott et al., 2005) (http://eutils.ncbi.nlm.nih.gov/entrez/query/static/esoap_help.html). The SOAP server enables users to write their own programs to automatically access the DBSubLoc server. A client with GUI interface is also provided for users who are not familiar with programming.

*To whom correspondence should be addressed.

OVERVIEW

The DBSubLoc SOAP server provides seven functions for remote users to access via HTTP protocol. Of the seven functions three are for querying and searching the database with Swiss-Prot ID (Boeckmann et al., 2003), Gene Ontology (GO) ID (Ashburner et al., 2000), DBSubLoc ID, protein name and BLAST (Altschul et al., 1997), one is for submitting users’ new sequences to the DBSubLoc database, while the other three are for protein subcellular location prediction with support vector machines (SVM) (Hua and Sun, 2001) method and PSORT (Nakai and Horton, 1999) program. More details of these functions are shown in Table 1. Supplementary tables describe the objects and returned values.

The SOAP server also comes with the WSDL (Web Services Description Language, http://www.w3.org/TR/wsdl) file at http://www.bioinfo.tsinghua.edu.cn/~tigerchen/SubLoc.wsdl, which makes it easy for users to program with a number of programming languages to access the SOAP server. For instance, with Perl version 5.8 and SOAP::Lite module, searching DBSubLoc database by protein name ‘p53’ would be as simple as the following example:

$wsdl = ‘http://www.bioinfo.tsinghua.edu.cn/~tigerchen/SubLoc.wsdl’;

$serv = SOAP::Lite->service($wsdl);

$res=$serv->name_search(‘p53’);

More detailed examples in several programming languages are available at http://www.bioinfo.tsinghua.edu.cn/~guotao/soapclient.html.

A statically precompiled binary of this open-source client with graphic user interface is available for download at http://www.bioinfo.tsinghua.edu.cn/~guotao/soapclient.html, together with screenshots, manual and FAQ. Executable binary files for Linux system and Microsoft Windows 2000/XP are also provided. Besides query by protein ID and name, the client software allows the user to load files containing multiple sequences in FASTA format for homology search with BLAST or for subcellular location prediction with SVM and PSORT. The result of searching and prediction could be saved to a user-defined text file.

IMPLEMENTATION

The SOAP server was implemented with Perl version 5.8. The server employs SOAP::Lite module to implement the SOAP
technology through HTTP protocol. MySQL version 4.0 serves as the database back end. The HTTP server daemon is Apache. The client is dependent on three Perl modules: Tk module for drawing GUI interface, SOAP::Lite module for communicating with the SOAP server and BioPerl module for parsing FASTA files.

ACKNOWLEDGEMENTS

We are grateful to anonymous reviewers for their constructive comments. This work was supported by Foundational Science Research Grant from the 973 project (2003CB715900), a National Nature Science Grant (No. 90303017, No. 90408019) and the 863 projects (2002AA234041, 2002AA231031).

Conflict of Interest: none declared.

<table>
<thead>
<tr>
<th>Method</th>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>name_search</td>
<td>Protein name, such as p53, actin and hemoglobin</td>
<td>Search DBSubLoc for proteins whose names contain the queried keyword</td>
</tr>
<tr>
<td>id_search</td>
<td>Protein ID for Swiss-Prot, DBSubLoc or GO. For example, 003478 (SWISSPROT id), 10021133 (DBSubLoc id) or 0005739 (GO id)</td>
<td>Search DBSubLoc for corresponding proteins according to queried id</td>
</tr>
<tr>
<td>blast_search</td>
<td>Protein sequence and E-value threshold</td>
<td>Search DBSubLoc by sequence homology search with BLAST</td>
</tr>
<tr>
<td>eu_predict</td>
<td>Protein sequence</td>
<td>Predict eukaryotic protein subcellular location with SVM method</td>
</tr>
<tr>
<td>pro_predict</td>
<td>Protein sequence</td>
<td>Predict prokaryotic protein subcellular location with SVM method</td>
</tr>
<tr>
<td>psort_predict</td>
<td>Protein sequence</td>
<td>Predict protein subcellular location with PSORT</td>
</tr>
<tr>
<td>feed_entry</td>
<td>Protein sequence and subcellular location information</td>
<td>Users could submit new protein sequences and subcellular location information to DBSubLoc via this function</td>
</tr>
</tbody>
</table>

REFERENCES


