ExPASy

水稻组: 吕桂云 张春秋
龚俊义 潘哲超

2008 年 1 月 12 日
第一部分：ExPASy简介
第二部分：数据库资源挖掘
第三部分：工具资源挖掘
第一部分：ExPASy简介

（1）数据库
（2）蛋白质分析工具
（3）配套基础性服务
（4）网络资源链接
第二部分：数据库信息资源挖掘
PROSITE: 一部宏大的蛋白质 motif 字典

PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns a profiles to identify them [More details / References / Disclaimer / Commercial users]. PROSITE is complemented by ProRule, a collection of rules based on profiles and patterns, which increases the discriminatory power profiles and patterns by providing additional information about functionally and/or structurally critical amino acids [More details].

Release 20.24, of 06-Dec-2007 (1500 documents)
Search in PROSITE for: leucine zipper

(Release 20.24, of 06-Dec-2007)

Enter search terms:
leucine zipper

[Image]

Choose bZIP

Basic-leucine zipper (bZIP) domain signature and profile

Description:

The bZIP superfamily [1,2] of eukaryotic DNA-binding transcription factors groups together proteins that contain a basic region mediated sequence-specific DNA-binding followed by a leucine zipper (see SP1/Cys2/Cys2) required for dimerization. bZIP domains usually bind palindromic 6 nucleotide site, but the specificity can be altered by interaction with accessory factor [3]. Several structures of bZIP have been solved [4]. The basic region and the leucine zipper form a contiguous helix where the four hydrophobic residues of the leucine zipper are oriented on one side. This conformation allows dimerization in parallel. Jun and Fos, two products of the Fos-related proteins 1-A and 1-B, and jun-B and jun-C, probable transcription factors which are highly similar to jun-A and -1, respectively, were shown to be involved in the regulation of the production of zymos protein during endoplasmic reticulum stress. The mammalian cAMP response element (CRE) binding proteins CREB, CREM, ATF-1, ATF-2, ATF-4, ATF-5, ATR, and CREF-1 are a group of transcriptional activators involved in the regulation of the production of zymos protein during endoplasmic reticulum stress. The mammalian cAMP response element (CRE) binding proteins CREB, CREM, ATF-1, ATF-2, ATF-4, ATF-5, ATR, and CREF-1 are a group of transcriptional activators involved in the regulation of the production of zymos protein during endoplasmic reticulum stress.

[Image]

bZIP 相关蛋白质保守结构序列的 dnalw 比对

bZIP 相关信息简单介绍和举例说明
bZIP 相关蛋白质在不同物种中的分类情况

bZIP 相关蛋白质各自的具体情况

bZIP 相关蛋白质的三维结构情况
输入待测蛋白的序列或登录识别号

分析得出的保守结构序列结果

对应的序列信息
PROSITE 数据资源非常丰富，适合于从感兴趣的保守结构入手，获取和分析某个或某一类具有相同共同保守结构序列的蛋白质信息。

数据库配套分析预测工具也很实用，它可以动用世界上最大最全的蛋白质 motif 信息资源来对待测序列进行全方位扫描，具有较高的可信度。

窗口命令输入形式具有多样化，适合于不同层次和不同要求的实验需要。
## ExPASy Proteomics Server

The ExPASy (Expert Protein Analysis System) proteomics server of the Swiss Institute of Bioinformatics (SIB) is dedicated to the analysis of protein sequences and structures as well as 2-D PAGE (Disclaimer / References).

### Databases
- UniProt Knowledgebase (Swiss-Prot and TrEMBL) - Protein knowledgebase
- PROSITE - Protein families and domains
- SWISS-2DPAGE - Two-dimensional polyacrylamide gel electrophoresis
- ENZYME - Enzyme nomenclature
- SWISS-MODEL Repository - Automatically generated protein models
- Links to many other molecular biology databases

### Tools and software packages
- Proteomics and sequence analysis tools
  - Identification and characterization (Adenta, FindMod, PepItam, ProSpine...)
  - Post-translational modification and topology prediction
  - Primary structure analysis
  - Secondary and tertiary structure tools (Swiss-PdbViewer...)
  - Alignment and Phylogenetic analysis
- **ImageMaster / Melanie** - Software for 2-D PAGE analysis
- MSight - Mass Spectrometry Imager
- Roche Applied Science's Biochemical Pathways

---

链接其它生物学信息资源
Protein related databases

Protein 3D structure related databases

Proteomics databases and links

Nucleotide and related databases

Carbohydrates resources

Species specific databases (Human/Vertebrates/Mitochondrion and chloroplast/Insects/Invertebrates/Plants/Fungi/Bacteria/Viruses and phages/

Human mutation databases/resources---)
GRAMENE: 粮食作物比较基因组学研究和QTL研究的最佳乐园
水稻最新基因组信息的大卖场

提供最新水稻基因组信息

美国孟山都水稻基因信息中心

提供最新水稻基因组注释信息
数据库生物信息手册涉及面相当广, 许多分门别类的生物学网络资源定能够成为我们日后实验的得力助手！
第三部分：分析工具资源挖掘
操作简单，适合于对核酸序列的初级蛋白质语言转化模拟。
操作简单，功能实用。
可以借助分析核酸序列中的外显子和内含子信息。
界面简洁，大大简化了搜索结果的表现形式，不会出现BLAST结果令人头晕目眩式的视觉效果。

配套序列分析工具尽管不是很全，但都是初级准备实验中非常适用的分析工具。
Motif Scan: 结构域搜索

Motif scanning means finding all known motifs that occur in a single protein sequence, select the collections of motifs to scan for, and the database of motifs (db description) to be used. The software documentation is available about the PROSITE and Pfam collections. For details of how to interpret the match scores, you should consult the PSSM, Pfam and InterPro for additional information.

Warning: The scan might take a few minutes, thus if your protein sequence is in database (see list), you should consider the query by protein form. The database of motifs (db description) provides a collection of tools that you might find useful.
搜索范围建立在包括PROSITE在内的很多数据库的基础之上，搜索结果具有很高的参考价值。

可以自己选择不同的搜索参数，进而可以有针对性对某一个或一类数据库进行搜索。

搜索结果不仅包含了最大可能性的保守性结构预测，而且还包括了与具有该结构的相关蛋白的比对信息和保守位点残基的频率信息，也即它具备SMART和MEME分析特点，还具备有序列基础性比对分析的多种功能。
序列扫描功能强大，结果选择项目多，适用于对未知序列的多种可能性motif预测。有效改变了其它相关分析工具搜索结果单一的不利状况。

搜索结果中还有对蛋白质初步细胞定位的信息。
### ChloroP v1.1 prediction results

#### Technical University of Denmark

<table>
<thead>
<tr>
<th>Name</th>
<th>Length</th>
<th>Score</th>
<th>cTP</th>
<th>CS—</th>
<th>cTP—</th>
</tr>
</thead>
<tbody>
<tr>
<td>P46644_AAT3_ARATH</td>
<td>449</td>
<td>0.513</td>
<td>Y</td>
<td>3.018</td>
<td>64</td>
</tr>
</tbody>
</table>

**ChloroP 叶绿体转移肽的预测**
MITOPROT: Prediction of mitochondrial targeting sequences

Enter your query sequence:
(Spaces, numbers, non-alphabetic characters will be automatically removed)

Input sequence length: 449 aa

VALUES OF COMPUTED PARAMETERS

Net charge of query sequence: +11
Analysed region: 27
Number of basic residues in targeting sequence: 4
Number of acidic residues in targeting sequence: 1
Cleavage site: 24
Cleaved sequence: MKTTHFSSSSSDRRIGALLRHL S

HYDROPHOBIC SCALE USED

<table>
<thead>
<tr>
<th></th>
<th>GES</th>
<th>KD</th>
<th>GVH1</th>
<th>ECS</th>
</tr>
</thead>
<tbody>
<tr>
<td>H17</td>
<td>1.406</td>
<td>1.324</td>
<td>0.122</td>
<td>0.565</td>
</tr>
<tr>
<td>MesoH</td>
<td>-0.366</td>
<td>0.460</td>
<td>-0.292</td>
<td>0.207</td>
</tr>
<tr>
<td>MuHd_075</td>
<td>21.659</td>
<td>10.879</td>
<td>5.981</td>
<td>4.317</td>
</tr>
<tr>
<td>MuHd_095</td>
<td>26.552</td>
<td>19.411</td>
<td>7.097</td>
<td>6.515</td>
</tr>
<tr>
<td>MuHd_100</td>
<td>29.312</td>
<td>20.372</td>
<td>7.359</td>
<td>7.920</td>
</tr>
<tr>
<td>Hmax_075</td>
<td>13.300</td>
<td>10.150</td>
<td>1.941</td>
<td>3.687</td>
</tr>
<tr>
<td>Hmax_095</td>
<td>7.438</td>
<td>10.800</td>
<td>2.483</td>
<td>3.560</td>
</tr>
<tr>
<td>Hmax_100</td>
<td>13.700</td>
<td>13.900</td>
<td>2.762</td>
<td>4.820</td>
</tr>
<tr>
<td>Hmax_105</td>
<td>1.700</td>
<td>11.900</td>
<td>0.578</td>
<td>4.060</td>
</tr>
</tbody>
</table>

PROBABILITY

of export to mitochondria: 0.5738
SignalP 3.0 Server

SignalP 3.0 server predicts the presence and location of N-terminal cleavage sites of different organisms: Gram-positive prokaryotes, Gram-negative prokaryotes, archaea, eukaryotes, and artificial neural networks and hidden Markov models.

View the version history of this server. All the previous versions.


Access the paper and supplementary information here.

Submit a file in FASTA format directly from your local disk:

Organism group
© Eukaryotes
© Gram-negative bacteria
© Gram-positive bacteria

Method
© Neural networks
© Hidden Markov models
© Both

Output format
© Standard
© Full
© Short (no graphics!)

Submit  Clear fields

SignalP 肽信号的预测

Paste a single sequence or several sequences in FASTA format into the:

>Q46644|AAT3_ARATH Aspartate aminotransferase, chloroplast - Arabidopsis thaliana (Mouse-ear cress).

Submit a file in FASTA format directly from your local disk:

Organism group
© Eukaryotes
© Gram-negative bacteria
© Gram-positive bacteria

Method
© Neural networks
© Hidden Markov models
© Both

Output format
© Standard
© Full
© Short (no graphics!)

Submit  Clear fields

SignalP 肽信号的预测

Predicted sequence:

MKTTHFSSSSSSRIGALLRKLHNSGSSD3NLSSLYAPSSTGSSTGGSSFVSHLVQAPEDPI

Secondary structure:

Predicted secondary structure:

SignalP-HMM prediction (set models): Q46644 ar17

Cleavage probability:

Non-secretory protein

Signal peptide probability: 0.002

Signal anchor probability: 0.002

METHODS

SignalP-HMM prediction (set models): Q46644 AR17

Cleavage probability:

Non-secretory protein

Signal peptide probability: 0.002

Signal anchor probability: 0.002

METHODS

SignalP-HMM prediction (set models): Q46644 AR17

Cleavage probability:

Non-secretory protein

Signal peptide probability: 0.002

Signal anchor probability: 0.002
TargetP 1.1 Server

TargetP 1.1 predicts the subcellular location of eukaryotic chloroplast transit peptide (cTP) and mitochondrial targeting signal (mTP). For submission, paste one or more sequences in FASTA format into the field below.

Submit a file in FASTA format directly from your local disk:

Organism group
- Non-plant
- Plant

Prediction scope
- Perform cleavage site prediction

Cutoffs
- No cutoffs; winner-takes-all (default)
- Specificity > 0.95 (predefined set of cutoffs that yielded this specificity)
- Specificity > 0.90 (predefined set of cutoffs that yielded this specificity)
- Define your own cutoffs (0.00 - 1.00): cTP: 0.00 mTP: 0.0

Submit Clear fields

TargetP 1.1 Server - prediction results
Technical University of Denmark

Reliability class
- Localization
- Presequence length

# targetp v1.1 prediction results
Number of query sequences: 1
Cleavage site predictions included.
Using PLANT networks.

<table>
<thead>
<tr>
<th>Name</th>
<th>Len</th>
<th>cTP</th>
<th>mTP</th>
<th>SP</th>
<th>other</th>
<th>Loc</th>
<th>RC</th>
<th>TP len</th>
</tr>
</thead>
<tbody>
<tr>
<td>P46644_AAT3_ARATH</td>
<td>449</td>
<td>0.545</td>
<td>0.310</td>
<td>0.036</td>
<td>0.149</td>
<td>C</td>
<td>4</td>
<td>64</td>
</tr>
</tbody>
</table>
NetNES 亮氨酸富集式细胞核输出信号分析预测

<table>
<thead>
<tr>
<th>Sequence</th>
<th>ANN</th>
<th>HMM</th>
<th>NES</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence-1</td>
<td>0.124</td>
<td>0.000</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-2</td>
<td>0.101</td>
<td>0.024</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-3</td>
<td>0.100</td>
<td>0.024</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-4</td>
<td>0.099</td>
<td>0.024</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-5</td>
<td>0.108</td>
<td>0.005</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-6</td>
<td>0.129</td>
<td>0.053</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-7</td>
<td>0.148</td>
<td>0.104</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-8</td>
<td>0.160</td>
<td>0.105</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-9</td>
<td>0.239</td>
<td>0.107</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-10</td>
<td>0.116</td>
<td>0.085</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-11</td>
<td>0.255</td>
<td>0.085</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-12</td>
<td>0.433</td>
<td>0.084</td>
<td>0.354</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-13</td>
<td>0.378</td>
<td>0.079</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-14</td>
<td>0.209</td>
<td>0.005</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-15</td>
<td>0.133</td>
<td>0.005</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-16</td>
<td>0.317</td>
<td>0.005</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-17</td>
<td>0.205</td>
<td>0.000</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-18</td>
<td>0.579</td>
<td>0.000</td>
<td>0.366</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-19</td>
<td>0.078</td>
<td>0.267</td>
<td>0.103</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-20</td>
<td>0.785</td>
<td>0.267</td>
<td>0.662</td>
<td>ret</td>
</tr>
<tr>
<td>Sequence-21</td>
<td>0.075</td>
<td>0.000</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-22</td>
<td>0.113</td>
<td>0.000</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-23</td>
<td>0.127</td>
<td>0.000</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-24</td>
<td>0.086</td>
<td>0.000</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-25</td>
<td>0.150</td>
<td>0.000</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-26</td>
<td>0.074</td>
<td>0.000</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-27</td>
<td>0.095</td>
<td>0.000</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-28</td>
<td>0.078</td>
<td>0.000</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-29</td>
<td>0.072</td>
<td>0.000</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Sequence-30</td>
<td>0.074</td>
<td>0.000</td>
<td>0.000</td>
<td>-</td>
</tr>
</tbody>
</table>

兼有信号预测和跨膜分析功能
Radar stands for Rapid Automatic Detection and Analysis of企業 networks. It is used to identify short composition biased as well as biased set architectures involving many different types of repeats in your query sequence.

**Radar Results**

Radar output: radar-20080111-07410935375062 output

---

<table>
<thead>
<tr>
<th>No. of Repeats</th>
<th>Total Score</th>
<th>Length</th>
<th>Diagonal</th>
<th>BW-From</th>
<th>BW-To</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>47.57</td>
<td>14</td>
<td>48</td>
<td>64</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>48 - 63 (21.24/24.41)</td>
<td>SSSLGFGELLgfQDLN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>307 - 320 (26.33/13.97)</td>
<td>SEMLGDGELL...QDLN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ProtParam 蛋白质基础信息的统计

输出结果中还包括半衰期、稳定性指数等信息。
蛋白质三级结构预测

- 同源建模法

瑞士生物信息研究所 SWISS-MODEL
丹麦技术大学生物序列分析中心 CPHmodels
比利时拿摩大学 ESyPred3D
英国癌症研究中心 3DJigsaw
SWISS-MODEL 蛋白质结构预测

预测过程
1. 从待测蛋白质序列出发，搜索蛋白质结构数据库（如 PDB, SWISS-PROT 等），得到许多相似序列（同源序列），选定其中一个（或几个）作为待测蛋白质序列的模板；
2. 待测蛋白质序列与选定的模板进行再次比对，插入各种可能的空位使两者的保守位置尽量对齐；
3. 建模：调整待测蛋白序列中主链各个原子的位置，产生与模板相同或相似的空间结构——待测蛋白质空间结构模型；
4. 利用能量最小化原理，使待测蛋白质侧链基团处于能量最小的位置。
SWISS-MODEL服务器可以反馈给用户以下结果选项，spdbv模式、普通模式和简短模式。这里，建议使用SWISS-MODEL默认的spdbv模式，这种模式可以用Swiss-PDBViewer程序打开，可以根据需要对结果进行多种操作。

注：预测结果将以邮件的形式发送给每一个客户。
致谢

- 感谢罗老师半年来的辛勤培育
- 感谢水稻组全体组员的通力协作
- 感谢caas07f1全班同学的关心与照顾