阅读日期：20190421

发表日期：20050211

发表杂志：Nucleic Acids Research

影响因子：

平均影响因子：

标签：

|  |  |
| --- | --- |
| 文章名 InterProScan: protein domains identifier | |
| 中文译名 | * InterProScan:蛋白质域标识符 |
| 网址 |  |
| 关键词 |  |
| 摘要 | InterProScan [E. M. Zdobnov and R. Apweiler (2001) Bioinformatics, 17, 847–848] is a tool that combines different protein signature recognition methods from the InterPro [N. J. Mulder, R. Apweiler, T. K. Attwood, A. Bairoch, A. Bateman, D. Binns, P. Bradley, P. Bork, P. Bucher, L. Cerutti et al. (2005) Nucleic Acids Res., 33, D201–D205] consortium member databases into one resource. At the time of writing there are 10 distinct publicly available databases in the application. Protein as well as DNA sequences can be analysed. A web-based version is accessible for academic and commercial organizations from the EBI (http:// www.ebi.ac.uk/InterProScan/). In addition, a standalone Perl version and a SOAP Web Service [J. Snell, D. Tidwell and P. Kulchenko (2001) Programming Web Services with SOAP, 1st edn. O’Reilly Publishers, Sebastopol, CA, http://www.w3.org/TR/soap/] are also available to the users. Various output formats are supported and include text tables, XML documents, as well as various graphs to help interpret the results. |
| 启发 | 这个就是结合多个数据库，预测蛋白质功能域的工具。 |
| 摘录 | 1  To search InterPro with a novel protein sequence, a tool, InterProScan (4), has been developed (<http://www.ebi.ac.uk/InterProScan> ) that combines the protein function recognition methods of the member databases of InterPro into one application.  本论文提出的工具是interPro版本的“升级版”  2  输入格式  首先填写你的邮箱，然后选择你想怎么看返回的数据。如果选“interactive”那么就返回浏览器，如果输入的是“email”则返回text dialog  第二步选择的是所需要的方法。若选膜蛋白，则单独勾选那一种方法。  第三个选择是针对DNA的。若输入的是DNA序列，则转换成蛋白质再寻找。还可以选最小收缩长度。  第四个选择是输入序列，默认输入的是protein的。还有help按钮可以使用。  最后的是“submit”“reset”按钮。  3网站上可以下载到这个工具，只有linux版本。没有Windows版和ios版的原因是“This is due to constraints in the various third-party binaries that InterProScan runs.”    现在的界面。<http://www.ebi.ac.uk/interpro/search/sequence-search>  4  输出格式  在开始预测前，它会计算所需要查询的序列，并将它的checksums和数据库已有蛋白质序列的进行比较。  这个数据库名叫“IPRMATCHES”它存有所有来自UniProt/Swiss-Prot and UniProt/TrEMBL的，匹配了一个或者多个的entries(词条)。若匹配结果不符，则同时启动分析程序。若相符，则返回已有条目。  5  最新版运行结果如下图。无“table”版本，可以导出多种形式。XML格式用于生产HTML视图。  运行结果存储7天（原文中是24小时）    6还有单机版，Stand alone版本。可以输入一个或者多个。单机版需要各种包。  （<http://www.ebi.ac.uk/interpro/interproscan.html> ）单机版可以在这里下载。    比如在这里就看到了各种各样的域。 |
| 相关论文 |  |

1，句子积累

Searching individually against each of these databases to get the most information is repetitive, time consuming and labour intensive.

2没有发现哪些功能域标识和药物相互作用，目前看到的功能域也是成段出现，没有成点出现。

3测试：输入药物作用靶点蛋白中这一条进行预测。

>3T6Q\_A

QKCIEKEVNKTYNCENLGLNEIPGTLPNSTECLEFSFNVLPTIQNTTFSRLINLTFLDLTRCQIYWIHED

TFQSQHRLDTLVLTANPLIFMAETALSGPKALKHLFFIQTGISSIDFIPLHNQKTLESLYLGSNHISSIK

LPKGFPTEKLKVLDFQNNAIHYLSKEDMSSLQQATNLSLNLNGNDIAGIEPGAFDSAVFQSLNFGGTQNL

LVIFKGLKNSTIQSLWLGTFEDMDDEDISPAVFEGLCEMSVESINLQKHYFFNISSNTFHCFSGLQELDL

TATHLSELPSGLVGLSTLKKLVLSANKFENLCQISASNFPSLTHLSIKGNTKRLELGTGCLENLENLREL

DLSHDDIETSDCCNLQLRNLSHLQSLNLSYNEPLSLKTEAFKECPQLELLDLAFTRLKVKDAQSPFQNLH

LLKVLNLSHSLLDISSEQLFDGLPALQHLNLQGNHFPKGNIQKTNSLQTLGRLEILVLSFCDLSSIDQHA

FTSLKMMNHVDLSHNRLTSSSIEALSHLKGIYLNLASNHISIILPSLLPILSQQRTINLRQNPLDCTCSN

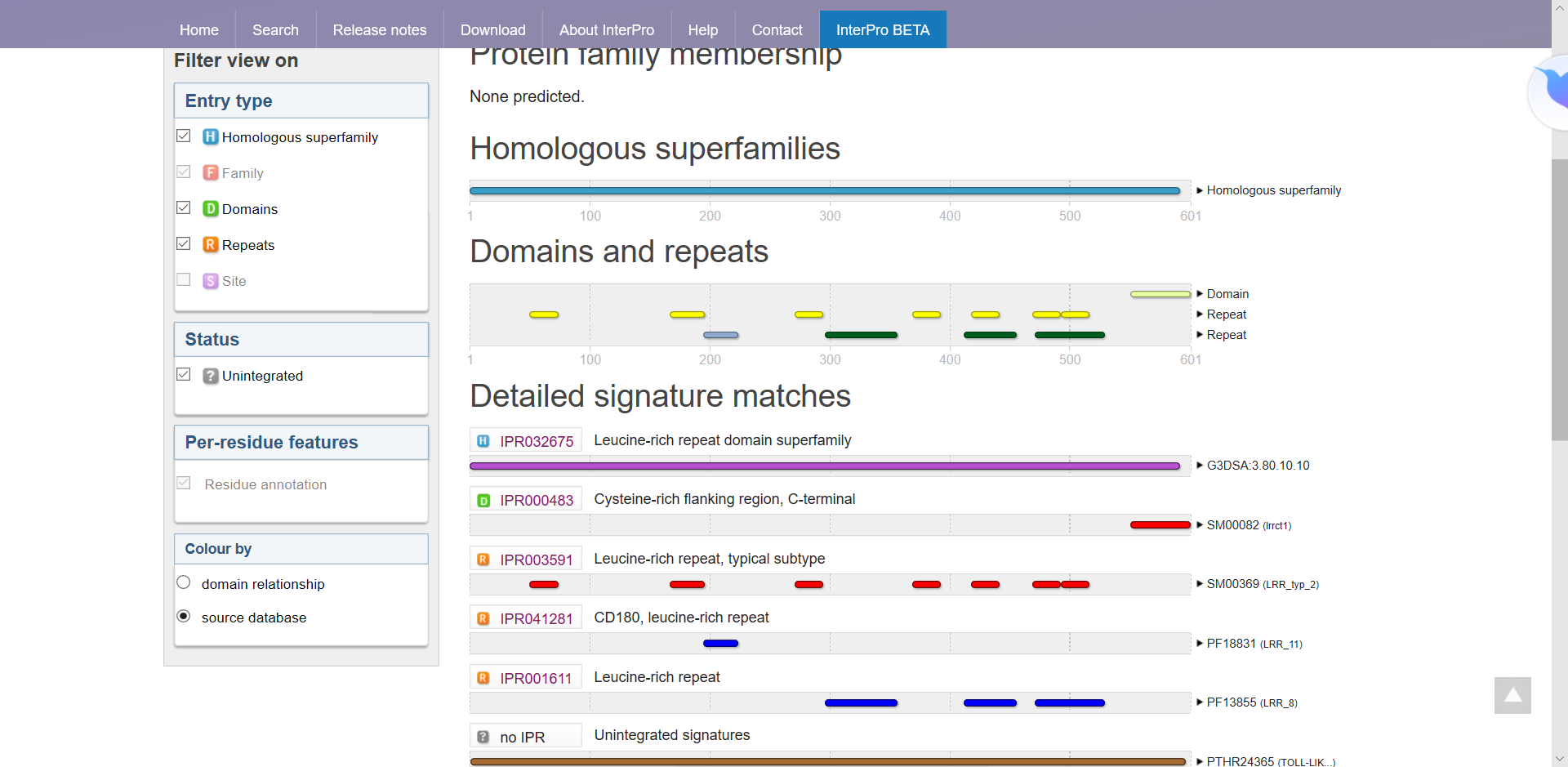
IYFLEWYKENMQKLEDTEDTLCENPPLLRGVRLSDVTLSCS

对应的功能位点：

>3T6Q\_A

QKCIEKEVNKTYNCENLGLNEIPGTLPNSTECLEFSFNVLPTIQNTTFSRLINLTFLDLTRCQIYWIHEDTFQSQHRLDTLVLTANPLIFMAETALSGPKALKHLFFIQTGISSIDFIPLHNQKTLESLYLGSNHISSIKLPKGFPTEKLKVLDFQNNAIHYLSKEDMSSLQQATNLSLNLNGNDIAGIEPGAFDSAVFQSLNFGGTQNLLVIFKGLKNSTIQSLWLGTFEDMDDEDISPAVFEGLCEMSVESINLQKHYFFNISSNTFHCFSGLQELDLTATHLSELPSGLVGLSTLKKLVLSANKFENLCQISASNFPSLTHLSIKGNTKRLELGTGCLENLENLRELDLSHDDIETSDCCNLQLRNLSHLQSLNLSYNEPLSLKTEAFKECPQLELLDLAFTRLKVKDAQSPFQNLHLLKVLNLSHSLLDISSEQLFDGLPALQHLNLQGNHFPKGNIQKTNSLQTLGRLEILVLSFCDLSSIDQHAFTSLKMMNHVDLSHNRLTSSSIEALSHLKGIYLNLASNHISIILPSLLPILSQQRTINLRQNPLDCTCSNIYFLEWYKENMQKLEDTEDTLCENPPLLRGVRLSDVTLSCS

0000000000000000000000000000000000000000000100000000000100000000110110000000000000000000000000000000000001010000110000000000000000000010000000000000000001010000000000000010010110000000000000000000111000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000010100000000000000000000011100000000000000000000000000000000000000000000000000000000000000000000000010000000000000000000000100000000000000000000000000010000000000000000000000011000000000000000000000101000000000000110000000000000000000000000000010000000000000000000000000000



具体有没有关系还是编程看吧，手工看没想到好的方案。



总之，看完这个感觉对当前课题没啥有用的帮助。不过知了前人做那个论文的蛋白质域是怎么来的了。