

ExPASy网络资源再挖掘

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



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).

[\[Announcements\]](#) [\[Job opening\]](#) [\[Partner Sites\]](#)

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 ([Aldente](#), [FindMod](#), [Popitam](#), [Phenyx](#), [pI/Mw](#), [ProtParam](#)...)
 - DNA -> Protein
 - Similarity searches ([BLAST](#)...)
 - Pattern and profile searches ([ScanProsite](#)...)
 - 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
- [Reche Applied Science's Biochemical Pathways](#)



Database of protein domains, families and functional sites

PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns and 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 of 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)

关键词类 表现特点类 分类学 其它符合度

e.g. PDOC00022, PS50089, SH3, zinc finger

Browse:

 add wildcard '*'

- by documentation entry
- by ProRule description
- by taxonomic scope
- by number of positive hits

Search in PROSITE for: leucine zipper

(Release 20.24, of 06-Dec-2007)

Enter search terms:

leucine zipper

Prefix and append wildcard '*' to words.

输入 leucine zipper

By default, this search engine searches for complete words only. If you did not find what you expected, and would like to do a substring match, you should perform a new search and select the "Match substrings" option.

选择 bZIP

Number of documents in PROSITE containing the search terms:

- PDOC51306 ASD1 and ASD2 domains profiles
- PDOC50188 B30.2/SPRY1 domain profile
- **PDOC00036 Basic-leucine zipper (bZIP) domain signature and profile**
- PDOC51229 DCUN1 domain profile
- PDOC51139 GTF2-like repeat profile
- PDOC50032 Kinase associated domain 1 (KAI1) profile
- PDOC00029 Leucine zipper pattern
- PDOC00347 Photosystem I psaA and psaB proteins signature
- PDOC00593 Sigma-54 factors family signatures and profile
- PDOC00991 TSC-22 / dip / bun family signature
- PDOC50994 Topaz/Def type signature and profile

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 mediating sequence-specific DNA-binding followed by a leucine zipper (see <PDOC00029>) required for dimerization. bZIP domains usually bind to a palindromic 6 nucleotide site, but the specificity can be altered by interaction with accessory factor [3].

Several structures of bZIP have been solved (see for example <PDB:1AN2>) [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 and it bends the helices so that the newly functional dimer forms a flexible fork where the basic domains, at the N-terminal open end, then interact with DNA. The two leucine zipper helices are separated by a DNA [4, 5].

This family is part of the larger family of transcription factors that includes the fos and jun families, the myc family, and metallothionein IA.

- Transcription factor 1 (TF1), also known as c-Jun, is a transcription factor that binds to the c-Jun response element (CRE) and metallothionein IA.
- Jun-B and Jun-D, probable transcription factors which are highly similar to jun/AP-1.
- The fos protein, a proto-oncogene that forms a non-covalent dimer with c-jun.
- The fos-related proteins fra-1, and fos B.
- Mammalian cAMP response element (CRE) binding proteins CREB, CREM, ATF-1, ATF-3, ATF-4, ATF-5, ATF-6 and LRF-1.
- Maize Opaque 2, a trans-acting transcriptional activator involved in the regulation of the production of zein proteins during endosperm development.
- Arabidopsis G-box binding factors GBF1 to GBF4, Parsley CPRF-1 to CPRF-3, Tobacco TAF-1 and wheat EMBP-1. All these proteins bind the G-box promoter elements of many plant genes.
- Drosophila protein Giant, which represses the expression of both the kruppel and knirps segmentation gap genes.
- Drosophila Box B binding factor 2 (BBF-2), a transcriptional activator that binds to fat body-specific enhancers of a subset of developmental genes.

bZIP 相关信息简单介绍和举例说明

4 proteins with a BTB, BZIP architecture:

O14867 (BACH1_HUMAN)



14 proteins with a KID, BZIP architecture:

P15337 (CREB1_RAT)



bZIP 相关蛋白质保守结构大致模型分类

7 proteins with a ZINC_FINGER architecture:

O93602 (ATF2_CHICK)



233 proteins with a BZIP architecture:

A11224 (CR3L2_DANRE)



CLUSTAL format alignment

```

ACA2_YEAST/427-482      --KRARLLERTRIAISKQKQKVAQIQKEFNEIKDENRIILKLNYYEKLISKFK-----
AP1_KLULA/51-114      TEAKDKRTAQRAAQRAPREERERKMKKEEDKVSQDESLNKQSELETKFLRNQVITNLSSELRK
AP1_SCHPO/76-139      QEPSSKKAQRAAQRAPREERERKMKKEEDKVSQDESLNKQSELETKFLRNQVITNLSSELRK
ATF1_BOVIN/212-263    QLRREILKLRREA REQRK KKEYVK ENRVAV ENQNKTIIEELKTLKED-----
ATF1_HUMAN/213-264   QLRREILKLRREA REQRK KKEYVK ENRVAV ENQNKTIIEELKTLKED-----
ATF1_MOUSE/211-262   QLRREILKLRREA REQRK KKEYVK ENRVAV ENQNKTIIEELKTLKED-----
ATF1_SCHPO/472-535   EEKRSFLERDQA LKQKQKQVLSHQAKVEFYGNENEISQAQVSALREEIVSLKTLLIAE
ATF21_SCHPO/267-330  DEKRRRFLERTRIA SKQKQKLVQNEKTAHIAICEQSKARILVSQLRREIVICLNKQLLAI
ATF2_CHICK/334-397  DEKRRRFLERTRIA SROKQKRVVYQS EKKAEDESSLNGQIQSEVTLRNEVAQLKQLLAI
ATF2_HUMAN/334-397  DEKRRRFLERTRIA SROKQKRVVYQS EKKAEDESSLNGQIQSEVTLRNEVAQLKQLLAI
ATF2_MOUSE/334-397  DEKRRRFLERTRIA SROKQKRVVYQS EKKAEDESSLNGQIQSEVTLRNEVAQLKQLLAI
ATF2_RAT/334-397    DEKRRRFLERTRIA SROKQKRVVYQS EKKAEDESSLNGQIQSEVTLRNEVAQLKQLLAI
ATF31_SCHPO/121-184  CTKQSPKLRRAA SROKQKRVVYQS EKKAEDESSLNGQIQSEVTLRNEVAQLKQLLAI
D1F3_BOVIN/86-144   D1F3_BOVIN/86-144  D1F3_BOVIN/86-144  D1F3_BOVIN/86-144
ATF3_HUMAN/86-144   ATF3_HUMAN/86-144  ATF3_HUMAN/86-144  ATF3_HUMAN/86-144
ATF3_MOUSE/86-144   ATF3_MOUSE/86-144  ATF3_MOUSE/86-144  ATF3_MOUSE/86-144
ATF3_RAT/86-144     ATF3_RAT/86-144    ATF3_RAT/86-144    ATF3_RAT/86-144
ATF4_BOVIN/275-338  VEKLLKMEQKTA TRYQKRAEQES NSECSE EKKNRE SEKADSLREIQYLRDLLEEF
ATF4_DANRE/266-329  LDKKLEMEQKTA TRYQKRAEQEA TGECKE EKKNEA KEKADSLAKEIQYKDLIEEY
ATF4_HUMAN/278-341  LDKKLEMEQKTA TRYQKRAEQEA TGECKE EKKNEA KEKADSLAKEIQYKDLIEEY
ATF4_MOUSE/276-339  LDKKLEMEQKTA TRYQKRAEQEA TGECKE EKKNEA KEKADSLAKEIQYKDLIEEY
ATF4_RAT/274-337    LDKKLEMEQKTA TRYQKRAEQEA TGECKE EKKNEA KEKADSLAKEIQYKDLIEEY
ATF5_HUMAN/208-271  GDRKQKRDQKSA LRYQKRAEGEA EGECQC EARNRE KRERAEVSEREIQYVKDLLIEY
ATF5_MOUSE/209-272  GDRKQKRDQKSA LRYQKRAEGEA EGECQC EARNRE KRERAEVSEREIQYVKDLLIEY
ATF5_RAT/207-270   GDRKQKRDQKSA LRYQKRAEGEA EGECQC EARNRE KRERAEVSEREIQYVKDLLIEY
ATF6A_HUMAN/306-369  VLLRQQNIEKRES CQSRRKKEYLQG EARLKAALSENEQKRENGTLERQDVEVSENRRI
ATF6B_HUMAN/325-388  VLLRQQNIEKRES CQSRRKKEYLQG EARLQAVLADNQQKRENAALRRRLKALLAENSSEI
ATF6B_MOUSE/322-385  VLLRQQNIEKRES CQSRRKKEYLQG EARLQAVLADNQQKRENAALRRRLKALLAENSSEI
ATF7_HUMAN/343-406  DERRORFLERTRIA SROKQKRVVYSS EKKAEETSONIQIQSEVTLRNEVAQLKQLLAI
ATF7_MOUSE/332-395  DERRORFLERTRIA SROKQKRVVYSS EKKAEETSONIQIQSEVTLRNEVAQIKQLI

```

bZIP 相关蛋白质保守结构序列的 clustalw 比对

Eukaryota - 2027

Metazoa - 961

Chordata - 766

Craniata - 736

Mus musculus (Mouse) - 162

Homo sapiens (Human) - 117

Danio rerio (Zebrafish) (*Brachydanio rerio*) - 78

Tetraodon nigroviridis (Green puffer) - 55

Rattus norvegicus (Rat) - 53

Xenopus laevis (African clawed frog) - 48

Bos taurus

bZIP 相关蛋白质在不同物种中的分类情况

Gallus gallus

Fugu rubripes (Japanese pufferfish) (*Takifugu rubripes*) - 21

Pongo pygmaeus (Orangutan) - 13

Ovis aries (Sheep) - 10

Sus scrofa (Pig) - 10

Oncorhynchus mykiss (Rainbow trout) (*Salmo gairdneri*) - 8

Ctenopharyngodon idella (Grass carp) - 5

Canis familiaris (Dog) - 4

Entry name	AC	Gene names	Description	Organisms
ACA2_YEAST	P40535	CST6, ACA2, YIL036W	ATF/CREB activator 2 (Chromosome stability protein CST6)	Saccharomyces cerevisiae (Baker's yeast)
AP1_KLULA	P56095	YAP1, KLLA0A01760g	AP-1-like transcription factor	Kluyveromyces lactis (Yeast) (Candida sphaerica)
AP1_SCHPO	Q01663	pap1, caf3, SPAC1783.07c	AP-1-like transcription factor (Caffeine resistance protein 3)	Schizosaccharomyces pombe (Fission yeast)
ATF1_BOVIN	Q08DA8	ATF1	Cyclic AMP-dependent transcription factor ATF-1 (Activating transcription factor 1)	Bos taurus (Bovine)
ATF1_HUMAN	P05117	ATF1	Cyclic AMP-dependent transcription factor ATF-1 (Activating transcription factor 1)	Homo sapiens (Human)
ATF1_MOUSE	P81269	Atf1	Cyclic AMP-dependent transcription factor ATF-1 (Activating transcription factor 1) (TCR-ATF1)	Mus musculus (Mouse)
ATF1_SCHPO	P52890	atf1, gad7, mts1, sss1, SPBC29B5.01	Transcription factor atf1 (Transcription factor mts1) (Protein sss1)	Schizosaccharomyces pombe (Fission yeast)
ATF21_SCHPO	P78962	atf21, SPBC2F12.09c	Transcription factor atf21	Schizosaccharomyces pombe (Fission yeast)

bZIP 相关蛋白质各自的具体情况

PDB	Swiss-Prot entries	Compound name (Submission date)
1A02	FOS_HUMAN (P01100) JUN_HUMAN (P05412) NFAC2_HUMAN (Q13469)	MOL_ID: 1; MOLECULE: DNA (5'- D (*TP*TP*GP*GP*AP*AP*AP*AP*TP*TP*GP*TP*TP*P P*TP*AP*G)-3'); CHAIN: A; ENGINEERED: YES; MOL_ID: 2; MOLECULE: DNA (5'- D (*AP*AP*CP*TP*AP*TP*GP*AP*AP*AP*CP*AP*AP*AP*TP P*TP*CP*C)-3'); CHAIN: B; ENGINEERED: YES; MOL_ID: 3; MOLECULE: NUCLEAR FACTOR OF ACTIVATED T CELL; CHAIN: N; SYNONYM: NFAT; ENGINEERED: YES; MOL_ID: 4; MOLECULE: AP-1 FRAGMENT FOS; CHAIN: F; FRAGMENT FOS; ENGINEERED: YES; MUTATION: YES; CHAIN: J; ENGINEERED: YES;
1DGC	GCN4_YEAST (P03069)	MOL_ID: 1; MOLECULE: PROTEIN (GCN4); CHAIN: A; MOL MOLECULE: DNA (5'- D (*TP*GP*GP*AP*GP*AP*TP*GP*AP*CP*GP*TP*CP*AP*TP P*CP*C)-3'); CHAIN: B; ENGINEERED: YES [3.00 Å] (15-JU
1DH3	CREB1_HOUSE (Q01147)	MOL_ID: 1; MOLECULE: TRANSCRIPTION FACTOR CREB; CHAIN: A FRAGMENT: RESIDUES 201-255; ENGINEERED: YES; MOL_ID: 2; MOLECULE: DNA (5'- D (*CP*CP*TP*TP*GP*GP*CP*TP*GP*AP*CP*GP*TP*CP*AP*GP*CP* *AP*G)-3'); CHAIN: B, D; ENGINEERED: YES [3.00 Å] (27-NOV-99)
1FOS	FOS_HUMAN (P01100) JUN_HUMAN (P05412)	MOL_ID: 1; MOLECULE: P55-C-FOS PROTO-ONCOGENE CHAIN: E, G; SYNONYM: CELLULAR ONCOGENE C-FOS, ENGINEERED: YES; MOL_ID: 2; MOLECULE: C-JUN PROT

bZIP 相关蛋白质的三维结构情况

PROSITE tools

Scan a sequence against PROSITE patterns and profiles - quick scan

- **ScanProsites** - advanced scan
- **PRATT** - allows to interactively generate conserved patterns from a series of unaligned proteins.
- **MyDomains - Image Creator** ^{new} - allows to generate domain figures.



(Output includes graphical view and feature detection)



Enter your sequence or a UniProtKB (Swiss-Prot or TrEMBL) ID or PDB AC [help]

P06731

输入待测蛋白的序列或登陆识别号

Scan Clear

exclude patterns with a high probability of occurrence

hits by profiles: [6 hits (by 1 profile) on 1 sequence]

Hits by P050835 IG_LIKE Ig-like domain profile:



Carcinoembryonic antigen-related cell adhesion molecule 5 precursor (Carcinoembryonic antigen) (CEA) (Meconium antigen 100) (CD66 antigen). *Homo sapiens* (Human)
Hits on PDB 3D structure: [1E07-A]

145 - 232: score = 10.39

PKPSSIShNSKPVEDKDAVAFTCEPET--QDATYLWVVN-----QSLPVspr1q1
SNGNRTLFLNVTNRNDTASVRCETQpVSARRSDSVILN

Predicted feature:

DISULFID 167 215

分析得出的保守结构序列结果。

240 - 315: score = 13.639

PTISPL--NTSYRSGENLNLSCHAAS-NPPAQYSWVFNQ-----TFQOST-----QEL
FIPNITVNSGSYTCQANNSDTGLNRTTVTT

Predicted feature:

DISULFID 259 299 By similarity

[condition: C-x*-C]

323 - 410: score = 9.337

PKPFIThNSNPVEDEDAVALTCEPEI--QNTTYLWVVN-----QSLPVspr1q1

Predicted feature:

DISULFID 345 393 By similarity

[condition: C-x*-C]

418 - 495: score = 13.021

PTISPSY--TYRPGVNLNLSLSCHAAS-NPPAQYSWLDIGN-----IQQHT-----QEL
FISNITEKNSGLYTCQANNsASGHSRTTVKTIIT

Predicted feature:

DISULFID 437 477 By similarity

[condition: C-x*-C]

501 - 588: score = 10.372

PKPSSIShNSKPVEDKDAVAFTCEPEA--QNTTYLWVVN-----QSLPVspr1q1
SNGNRTLFLNVTNRNDARAYVCGIQNaVSANRSDPVTLN

Predicted feature:

DISULFID 523 571 By similarity

[condition: C-x*-C]

593 - 675: score = 12.604

PDTP IISpPDSSYLSGANLNLSCHSAS-NPSPQYSWRINGI-----PQOHT-----QV
LFIARIKIPNNNGTYACFVSNlactGRNNSIVKSIIVS

Predicted feature:

DISULFID 615 655 By similarity

[condition: C-x*-C]



ScanProsites Results Viewer

This view shows ScanProsites results together with ProRule-based predicted intra-domain features (help).

hits for all PROSITE (release 54.1) and UniProtKB (Swiss-Prot (release 54.1))

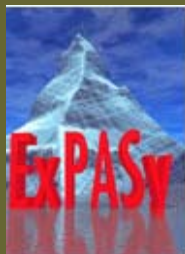
Found: 6 hits in 1 s

对应的序列信息

P06731 CEAM5_HUMAN (702 aa)
Carcinoembryonic antigen-related cell adhesion molecule 5 precursor (Carcinoembryonic antigen) (CEA) (Meconium antigen 100) (CD66 antigen). *Homo sapiens* (Human)

PKPSSIShNSKPVEDKDAVAFTCEPET--QDATYLWVVN-----QSLPVspr1q1
SNGNRTLFLNVTNRNDTASVRCETQpVSARRSDSVILN
PTISPL--NTSYRSGENLNLSCHAAS-NPPAQYSWVFNQ-----TFQOST-----QEL
FIPNITVNSGSYTCQANNSDTGLNRTTVTT
PKPFIThNSNPVEDEDAVALTCEPEI--QNTTYLWVVN-----QSLPVspr1q1
PTISPSY--TYRPGVNLNLSLSCHAAS-NPPAQYSWLDIGN-----IQQHT-----QEL
FISNITEKNSGLYTCQANNsASGHSRTTVKTIIT
PKPSSIShNSKPVEDKDAVAFTCEPEA--QNTTYLWVVN-----QSLPVspr1q1
SNGNRTLFLNVTNRNDARAYVCGIQNaVSANRSDPVTLN
PDTP IISpPDSSYLSGANLNLSCHSAS-NPSPQYSWRINGI-----PQOHT-----QV
LFIARIKIPNNNGTYACFVSNlactGRNNSIVKSIIVS

- 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).

[\[Announcements\]](#) [\[Job opening\]](#) [\[Mirror Sites\]](#)

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 (Aldente, FindMod, Popitam, ...)
 - **链接其它生物学信息资源**
 - Similarity searches (BLAST, ...)
 - Pattern and profile searches (ScanProsite...)
 - 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**



ExPASy Life Science Directory

(formerly known as [Amos](#)' WWW links page)

Notes:

- 1) The URL for this page is <http://www.expasy.org/links.html>
- 2) If you would like to submit a specific link or to notify us of a modified link, please [send us an email](#), but remember that we reserve the right to choose the links we want to include !
- 3) Links to protein sequence, 3D structure and 2D-gel analytical tools are provided on ExPASy's [Proteomics tools](#) page.

- **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/Vir
uses 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

The image shows a screenshot of the GRAMENE website. The page features a navigation bar with links for Search, Genomes, Species, Download, Resources, About, and Help. A search box is located in the top right corner. The main content area is titled "Quick Start" and lists several key resources:

- GENOMES:** Browse assembled genomes for *Oryza Sativa indica* and *Arabidopsis thaliana*; Look for *rice/maze synteny*; Narrow your search by alignment with BLAST; search by Gene Ontology.
- PROTEINS:** Search by PFam or ProSite or Browse by Gene Ontology using GO Slim.
- COMPARATIVE MAPS:** Browse genetic or physical maps for Wild Rice, Wheat, Barley, Oats, Sorghum, and other grasses, or use the Comparative maps of different types and species. View map detail information.
- MARKERS:** Search for Genetic markers (RFLPs, SSRs, etc.), DNA Probes (Primers, Overgos, etc.), Genomic Regions (Clones, FPContigs, etc.), and Sequences (GSSs, ESTs, etc.). Search by species such as *Sorghum*, by type such as *SSR Primer Pair*, or by species and type such as *Rice SSR*. Use the Simple Sequence Repeat Identification Tool (SSRIT).
- TRAITS:** Search the Genes or QTL database for important phenotypes, QTL, Maize QTL. Don't forget to explore traits in Ontologies.
- GENETIC DIVERSITY:** Search for SNP and SSR allelic variation on loci of *rice*, *maize*, and *wheat* germplasm.
- BIOCHEMICAL PATHWAYS:** Search for a gene, protein, or pathway. Search for known reactions (exempli gratia *starch biosynthesis*) or get an overview.
- LITERATURE:** Search literature for your friends and topics of interest.

Four red callouts with arrows point to specific sections of the page:

- 比较基因组学资源** (Comparative Genomics Resources) points to the GENOMES section.
- 分子标记资源开发** (Molecular Marker Resource Development) points to the MARKERS section.
- 生物化学途径** (Biochemical Pathways) points to the BIOCHEMICAL PATHWAYS section.
- 文献** (Literature) points to the LITERATURE section.

On the left side of the page, there is a sidebar with the following sections:

- Resource for Comparative Grass Genomics**
- Release # 26** (September 2007)
- News**
 - Gramene Receives NSF Grant funding. See NSF Award announcement
 - Gramene uses Google Outreach Calendar
 - Sept 2007, V 26 release notes.
 - October 15-17, 2007 - 5th International Symposium of Rice Functional Genomics in Tsukuba, Japan
 - September/October Gramene Newsletter
- Have Questions...?**
 - Quick Search Help
 - Tutorials - See what you can do and how to do it.

水稻最新基因组信息的大卖场

RGP
Rice Genome Research Program

Search: (Include) | JAPANESE | Site Map

Topics:

[The Whole-Genome Annotations for Rice is now open to the public.](#)

[An obituary to a pioneering molecular biologist and advocate of rice genome research.](#)

[Redistribution of genomic sequence \(draft 2\) of the cultivar Nipponbare of Oryza sativa ssp. japonica, assisted with RAP annotation.](#)

2 April 2007:
New datasets (drafts) are now added to the published rice genome annotation site.

28 February 2007:
IRGSP Website is the Current Web contents of Thomson Scientific.

9 September 2006:
The rice-based sequence of the rice genome feature 426,700,000 is selected as one of New Hot Papers by ISI.

1 September 2006:

Oryzabase
Integrated Rice Science Database

Search: All Sections | |

Strains | Wild rices (collections) | Mutants | Development | Genes | References | Distribution

Linkage Maps | Physical Maps | Comparative Maps | Basic | Dna seq. | BLAST | Organelles | Tools & Protocols | Download

Database contents:

NBRP Strains	15,842 entries	Strains	12,426 entries
	4,904 entries	Trait genes	3,258 entries

Hot News !!

RGN

Featured Links:

- RGRC
- IRGSP
- RAP
- KOME
- Gramene
- TIGR Rice
- IRRI-CRC

提供最新水稻基因组信息

rice-research.org

Monsanto launched the rice-research.org public database in June of 2000 to provide access to the Monsanto draft rice genome sequences to public researchers. Almost 800

years

At the end of 2002, the International Rice Genome Sequencing Project (IRGSP) completed the Japanese rice variety Nipponbare genome sequence at Phase II or at finished quality. More than 500 Mb of these sequences are available in public databases and the final finished sequence will follow soon afterwards - please refer to the IRGSP sites for updates (e.g. <http://rgp.dna.affrc.go.jp/cgi-bin/status@seqolab.jp> or <http://www.tigr.org/tdb/e31/osa1/>).

美国孟山都水稻基因信息中心

TIGR Rice Genome Annotation

Home | Site Map | Download | FAQ | Links | Genome Browser

Project Overview

What's New

Pseudomolecules

Rice Genome Facts

Gene Nomenclature

Genome Annotation

...on project and provides sequence and annotation data for the rice genome. Through these pages, you can view, search, and download our annotation of the rice genome. All of the data is freely available and comments on the data or improvements to these pages are welcome.

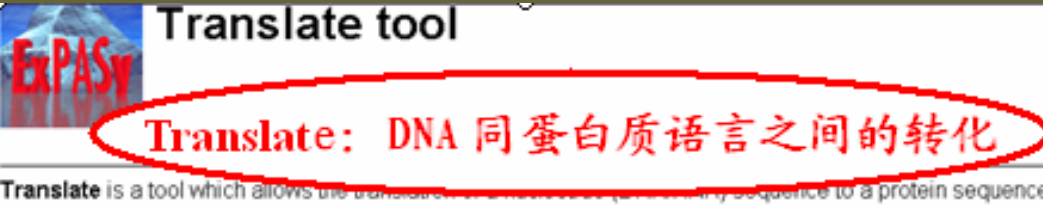
The major features of our pages are the sequence and annotation of the 12 rice chromosomes and the Genome Browser that provides an integrated display of annotation data.

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

DNA → Protein



Translate Tool - Results of translation

```
>gi|162462052|ref|NM_001111496.1| Zea mays hemoglobin (LOC541815), mRNA
GGTCACACCCCAACCTCCACACTGTA AAAAAGAGCAGCGGGAACGTGGCTGCATCCATCCATCCATTTCC
CAATCCCAATCCCAATCCCAACCCAGTGTCCAGTGCTCGGGGAACCGACACAGCTCCTCAGCAGAGTAGCCA
GCACGACAAGCCCGATCAGCAGACAGCAGGCATGGCACTCGCGGAGGCCAGCAGCCGGCGGTGGTCTTC
GGCGAGGAGCAGGAGGCCGCTGGTGCTCAAGTCGTGGGCGGTCATGAAGAAAGGACGCCGCCAACCTGGGGC
TCCGCTTCTTTCTCAAAGTCTTCGAGATCGCGCCGTCGGCGGAAGCAGATGTTCTCGTTCCTGCGGCACTC
CGACGTGCGCGTGGAGAAGAACCCCAAGCTCAAGACGCAGCCATGTCGCTTTCGTATGACCTGCGGAG
CGCGCGCGCAGCTTCGCAAGGCCGGGAAGGTCAACCGTGAGGGAGACCCAGCTCAAGAGGCTGGCGCCA
GGCACTTGAGGTAGCGCGTCCGAGATGGACACTTCGAGGTGACGGGGTTCGCGCTGCTTGAGACGATCAA
GGAGGGCGCTCCCGCTGACATGTGGAGGCTCGAGATGAAGAAAGCTGGGCCGAGGCGCTACAGCCAGCTG
GTGGCGCCATCAAGCGGAGATGAAGCCGATGCCATGATGTCGCGATTCCGACCACTGTTTAAACCCAT
GACGCAGCGCGCTCACAGATGTCGCGTGTGGCTTTCGCTTTAGCAATTTCTCTCGGAGGGAGCGTGTGA
TTGTTATCTTGATCGAGAGCCGCTGTGTGCTGCGCTTTGCTTCTGTGATTATATACTACTGAATAAAGA
TGTAGCGTTCTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

Please select one of the following frames:

```
5'3' Frame 1
XXXXXXXXXXRSHTSHCKKEQRNVRASIHPSISNPNPNTSVQC SGNRHSSSAE Stop
ARSADSRHGTRGRRRRRGGGLRRGAGGAGAQQVGRHEEGRRQPGPPLLSQGLRDRAVGE
VPARLRRRAAGEEPQAQDARHVLRHLDRGGGAASQGREGHRGHDHAQEAGRHALEVRR
LRGDGVRAA Stop DDQGGAPR Stop HVEPRDEESLGRGLQPAGGGHQAGDEARCLVVAIAT
Stop P Iet TQRRRHRCPVWSCALAI SLWRERVLLS CDREPVCCLCFL Stop LYSY Stop IK Iet S
SSKKKKKKKKK

5'3' Frame 2
XXXXXXXXXXGHTQPPTVKKSSGTCVHPSIHPFPIPIPIPPVSSARGTDTAPQSSQH
QQTAG Iet ALAEADDGAVVFGEEQALV LKSWAV Iet KKDAANLGLRFFLKVF EIAPSAK
FSFLRDS DVPLEKNPKLKTHA Iet SVFV Iet TCEAAAQLRKAGKVTVRETT LKRLGATHL
VADGHFEVTFGALLETIKEALPAD Iet WSLE Iet KKAWAEAYSQ LVAAIKRE Iet KPDA St
Stop WRLRPVFN P Stop RSAVTDVPCGLAL Stop QFLSGGSVYCYLV IE SLCAAFASCDYIA
Stop RC SVLQKKKKKKKKK

NLP L Stop KRAAERACIHP SIHFQS QSQSHQCPVLGEPTQL LSRVAS
TTARWSSARSRRRWCSSRGPS Stop RRTPTWASASF SRSSRSRRR
TPSSRRRTPCPSSS Stop PARRRRSFAR PGRSP Stop GRP RSRGWAPRT
GTASQ Iet DTSR Stop RGS RCLRRSRRRSP LTCGASR Stop RKPGRPTASWWRPSSGR Stop
Iet PSSGDCDQLTHDAAPSQ Iet SRVLRFSNFSLEGACI VIL Stop SRACVLP LLLV I
LNKDVAFFKKKKKKKKK

5'3' Frame 1
FFFFFFFFFFFFLKNATSLFSSYI ITRSKGSLTQALDHKITIHAPSREKLLKRKTTRD ICD
AVKHWSSQSPLLIGLHLPDGRHQLAVGLPGFLHLEAPHVSGERLLDR LKQREPRHL
CDAVPQVRGAQPLERGLPHGDLPLGLAKLRRLRAGHDEDGHGVRLELG VLLQRHVGVVAQ
LLRRRRDLEDLEKEAEAQVGVLLHDGPRLEHQRLLLLAEDHRAVVGLRECHACCLLI
LATLLRSCVGS PSTGHWWDWDWDWKW Iet DGW Iet HARSAALFYSGRLGV TXXXXXXXXX

5'3' Frame 2
FFFFFFFFFFFF Stop RTLHLYSVAI Stop SQEAKAAHRLSITR Stop QYTLPPERNC Stop SAR
SVTALRHGLNTGRNRHY Stop ASGFISRL Iet AATS WL Stop ASAQAFFISRLH Iet SAGSA
SSSANPVTSKCP SATPYLKC VAPSLLSVVS LVTFFPALRSCAAASQV Iet TKTD Iet ACV
GFFSSGTS ESRNENICFADGAI SKTLRKKRRPRLAASF Iet TAHDLS T SASCSSPKT
SSASASA Iet PAVC Stop SGLSCWLLC Stop GAVSVPRALDTGGIGIG IGNW Iet DGCTHV
FTVGGWV Stop PXXXXXXXXX

5'3' Frame 3
FFFFFFFFFFFFEERYIFIQ Stop LYNHKKQRQHTGSR S QDNNTRS LQREIAKAQDHTGHL S
RCV Iet G Stop TLVAIAT TRHRASSPA Stop WPPPGCRPRRLSSSRG STCQRGAP Stop S
ARTPSPRSVHLRRRTSSAWRPAS Stop AWSPSR Stop PSRPEC AAPP RRS Stop RRRTWV
```

Output format: Verbose ("Met", "Stop", spaces between residues) or

■ 操作简单,适合于对核酸序列的初级蛋白质语言转化模拟.

Wise: 外显子和内含子分析

The Wise2 form compares a protein sequence to a genomic DNA sequence, allowing for introns and frameshifting errors.

The model parameters which have been chosen for you are as follows:

- Human gene parameters
- Local start/end in the protein
- 6:23 Algorithm

This is the simple Wise2 form, to change these parameters you need to use the [advanced form](#).

[Download Software](#)

Your Email:

Results:

Output for alignments: Parameters Pretty alignment

Output for gene predictions: Gene structure Translation cDNA EMBL feature table format

Sequence 1: paste a Protein Sequence in FASTA format OR upload a file: Help

```
>P31946|1433B_HUMAN 14-3-3 protein beta/alpha - Homo sapiens (Human).
MTMDKSELVQKAKLAEQAERYDDMAAANKAVTEQGHEL SNEERNLLSVAYKNVVGARRSS
WRVISSIEQKTERNEKKQMGKEYREKIEAELQD ICNDVLELLDKYLIPNATQPESKVFY
LKMKGDYFRYLSEVASGDNKQTTVSNSQQAYQEA FEISKEMQPTHP IRLGLALNFSVFY
YEILNSPEKACSLAKTAFDEAIAELDTLNEESYKDSL IIMQLLRDNLTLWTSENQGDGEGD
AGEGEN
```

Seq. 1 Upload a file: 浏览...

Sequence 2: paste a DNA Sequence in FASTA format OR upload a file: Help

```
>gi|2791551|emb|AL008725.1| Human DNA sequence from clone RP1-148E22
GATCACGCTATTGCACCCCAGCCTGGCAAAAAGAGCGAGACTCCATCTCAAAAAACACACACACAAACAA
A
AAACAAACCACAAAGAGCATTTCAAATACTTTTGCTAAAAATCAACAGCTCTTCCCTTAGGTATCCCACAG
T
ACTGGACTACTCTCATTTCCTTCAGACCTCTTCTCAAATGTCAGCTACTCAGCGATACCCTGACCACCC
T
ATCCAAAATAGTAGCTCTCATGCCCAACCCCAATTCTCTGTCTCTTCTTACTGCTTGTGCATAATT
A
TGTAATTTTTTTGTGCTAGAACAATGCTTCATGAGAGCAGGGGCTGTGTCTGTTTTGTACACTGCTGC
```

Seq. 2 Upload a file: 浏览... Run Reset

■ 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.

相似性序列搜索

The screenshot shows the SEQUEROME web interface. Red boxes and arrows highlight several key features:

- 保守基序分析** (Conserved Motif Analysis): Points to the 'protparam' button in the 'Commands' menu.
- 基础统计分析** (Basic Statistical Analysis): Points to the 'Protein Stats' button in the 'Commands' menu.
- 表现形式分析** (Expression Analysis): Points to the 'Protein Pattern' button in the 'Commands' menu.
- 二维结构预测分析** (2D Structure Prediction Analysis): Points to the '2° Structure Prediction' button in the 'Commands' menu.
- 核酸定位信号肽预测** (Nucleic Acid Localization Signal Peptide Prediction): Points to the 'NLS Predict' button in the 'Commands' menu.
- 输入序列** (Input Sequence): Points to the text input field for the sequence.
- 序列比对键** (Sequence Alignment Key): Points to the 'BLAST' button.
- 序列类型选择** (Sequence Type Selection): Points to the radio buttons for 'DNA', 'RNA', and 'Protein'.
- 结果输出框** (Result Output Box): Points to the table of search results.

The search results table is as follows:

ID	Seq	Length	E-value
<input type="radio"/> gid: 60653657	View Seq	286	3e-76
<input type="radio"/> gid: 4504345	View Seq	286	3e-76
<input type="radio"/> gid: 8101647	View Seq	285	6e-76
<input type="radio"/> gid: 22671717	View Seq	285	6e-76
<input type="radio"/> gid: 62898345	View Seq	285	1e-75
<input type="radio"/> gid: 27574244 - pdbid: 1O1N	View Seq	285	1e-75
<input type="radio"/> gid: 9256890 - pdbid: 1C7D	View Seq	285	1e-75

- 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 protein sequence, select the collections of motifs to scan for, a [documentation](#) is available about the Prosite and Pfam collections. [Prosite documentation](#) deals with the interpretation of the match scores. You should check [ExPASy](#), [Pfam](#) and [InterPro](#) for additional information.

Warning: The scan might take a few minutes, thus if your protein is large, the [Query by Protein](#) form is more suitable. It 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.

Functional site prediction

Protein sequence
Enter SWISS-PROT/TrEMBL identifier or accession number:

Or paste the sequence (Single letter code sequence only or FASTA format):

```
>gi|28573655|ref|NP_611806.3| pita [Drosophila melanogaster]
MAAKLEKREAMLTEKRVCRFLTEQKLASIFENPRLVITANLPLQINAITAIEVYAGDG
NPGHICLECHLLFEBYCFRQMCRAETLLRQYVSTGNVPSFLEKPRAPRTHVASKLLLV
VPKATAEPSKTPKLLNTRAKSSQVIEKLVLESANVTPRTVAGSPVPRRSHATELK
VDNNGQLSHDDVQSELEDMASELEKEFPDIPQKASPVKPKVLNKSIRILNKOPAAPVEP
RLATPKVKRRDSSGNVAIVTEVLSDPLDQDDPTNAEKVATDVFFCPDCERSFPLQQL
LEIHLNHTRRSFCQLLCEKSFYSKYDLAKHNFVHTGERPFKCAICSKAFTRKALLNRH
ERTHTDVPFKICVYCEKFFLSRQENKHAERHQKRRPFCQGVCTKSAFAFRQGLERHETVH
```

Context information
Species
select from list
Homo sapiens
or type in manually:

Cell compartment (one or several):
not specified
extracellular
nucleus
cytosol

Submit Reset Form

Please be patient! The ELM server collects the SMART/FFam information through the SMART public queue.

Release Notes
The current release of the ELM database (Dec 2007) contains 132 ELMs.

序列输入区

选择研究物种

选择不同的结果

The ELM server
... known ELM instances and predictions in sequences similar to ELM instance sequences, where the motif is positionally conserved, are identified and displayed (see ELM instance mapper).
Users are encouraged to contribute information.



多样化的分析结果

结构注释
信号分析

Filtering summary

No user supplied cellular location.
User supplied taxon: Homo sapiens

(The ELM is listed as filtered when all its matching instances have been filtered out)

	Elms	Instances
retained	29	88
Species filtered	1	7
Smart filtered	10	0
Cellular location filtered	0	0
all found (before filtering)	40	162

Query sequence:
 gi|28573655|ref|NP_611806.3| pita [Drosophila melanogaster]
 MAAKLEKREAMLTEKRVCRFLTEQKLASIFENPRLVITANLPLQINAITAIEVYAGDG
 NPGHICLECHLLFEBYCFRQMCRAETLLRQYVSTGNVPSFLEKPRAPRTHVASKLLLV
 VPKATAEPSKTPKLLNTRAKSSQVIEKLVLESANVTPRTVAGSPVPRRSHATELK
 VDNNGQLSHDDVQSELEDMASELEKEFPDIPQKASPVKPKVLNKSIRILNKOPAAPVEP
 RLATPKVKRRDSSGNVAIVTEVLSDPLDQDDPTNAEKVATDVFFCPDCERSFPLQQL
 LEIHLNHTRRSFCQLLCEKSFYSKYDLAKHNFVHTGERPFKCAICSKAFTRKALLNRH
 ERTHTDVPFKICVYCEKFFLSRQENKHAERHQKRRPFCQGVCTKSAFAFRQGLERHETVH
 PRQKTRTFPPKDAQSQRVTRSDAQRITQKTKGKRGKPRHKKLNRKSRSAK

Results of ELM motif search after globular domain filtering and context filtering.

Matches falling inside SMART/FFAM domains are excluded from this list.

Elm Name	Instances (Matched Sequence)	Positions	Pro-Blast Instance Mapping	Elm Description	Cellular Location	Protein
CLV_NDR_NDR_1	RRS GRG RRS	172-174 633-635 679-681	-	N-Arg dibasic convertase (NDR) cleavage site (Lys-Arg)-His		RR1 (JALM)
CLV_PCSK_PCIETZ_1	KRD KRP KRA	249-250 395 451-453	-	MEC1/NEC2 cleavage site (Lys-Arg)-His	Cellular membrane, extracellular, Golgi apparatus	KR1 (JALM), L7V7
CLV_PCSK_SKI_1	KLLLV KLLLN RLATP KATD RNLJN	116-120 133-137 241-245 280-284 674-678	-	Subtilisin/kexin isozyme-1 (SKI) cleavage site (RQ)-G (hydrophobic)(L7V7)-G	endoplasmic reticulum, endoplasmic reticulum lumen, Golgi apparatus	RR1 (JALM), L7V7
US_CYCLIN_1	KLLLV KLLW KLLL	116-120 117-121 133-136	-	Substrate recognition site that interacts with cyclin and thereby increases phosphorylation by cyclin/cdk complexes. Predicted protein should have the MOD_CDK site. Also used by cyclin inhibitors	nucleus, cytosol	RR1 (JALM), FYLWMP
US_EVH1_1	FFDP	207-211	-	proline-rich motif binding to signal transduction class I EVH1	cytosol, cell membrane	FP (JALM)

详细结构域注释

- 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 叶绿体转移肽的预测

The ChloroP server predicts chloroplast transit peptides (cTPs) in protein sequences. A related service is available for predicting chloroplast transit peptides.

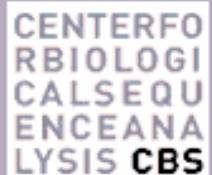
[Instructions](#)[Output format](#)[Data sets](#)[Abstract](#)

SUBMISSION

Paste a single sequence or several sequences in [FASTA](#) format into the field below:

```
>P46644|AAT3_ARATH Aspartate aminotransferase, chloroplast -  
Arabidopsis thaliana (Mouse-ear cress).  
MKITTFSSSSSSDRRIGALLRHLNSGSDSDNLSSLYASPTSGGTGGSTVFSHLVQAPEDPI
```

Submit a file in [FASTA](#) format directly from your local disk:

 Detailed output 

ChloroP 1.1 Server - prediction results

Technical University of Denmark

```
### chlorop v1.1 prediction results #####  
Number of query sequences: 1
```

Name	Length	Score	cTP	CS- score	cTP- length
P46644_AAT3_ARATH	449	0.513	Y	3.018	64

MITOPROT 线粒体接合肽信号的预测

Enter your query sequence:

(Spaces, numbers, non-alphabetic characters will be automatically removed)

MKTTTHFSSSSSSDRRIGALLRHLNSGSDSDNLSSLYASPTSGGTGGSVF
 VQAPEDPI
 LGVTVAYNKDPSPVKLN LGVGAYRTEEGKPLVLNWRKAEQQLINDRTR
 YLPITVGLV
 EFNKLSAKLILGADSPAIRENRITTV ECLSGTGSLRVGG EFLAKHYHQTI
 QPTWGN
 HPKIFTLAGLTVKTYRYYPATRGLNFQGLLEDLGAAAPGSIVLLHACAHF
 GVDPTIQ
 QWEQIRKLMRSKGLMPFFDSAYQGFASGLD TDAKPIRMFVADGG ECL

Input sequence length : 449 aa

Sequence name (optional):

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 : MKTTTHFSSSSSSDRRIGALLRHL

HYDROPHOBIC SCALE USED

	GES	KD	GVH1	BCS
H17	1.406	1.324	0.122	0.565
MesoH	-0.366	0.460	-0.292	0.207
NuHd_075	21.659	10.879	5.981	4.317
NuHd_095	26.552	19.411	7.097	6.515
NuHd_100	29.312	20.372	7.359	7.920
NuHd_105	30.297	21.562	7.424	9.028
Hmax_075	13.300	10.150	1.941	3.687
Hmax_095	7.438	10.800	2.483	3.560
Hmax_100	13.700	13.900	2.762	4.820
Hmax_105	1.700	11.900	0.578	4.060

PROBABILITY

of export to mitochondria: 0.5738

SignalP 3.0 Server

SignalP 3.0 server predicts the presence and location of secretory proteins in different organisms: Gram-positive prokaryotes, Gram-negative prokaryotes, and eukaryotes. It also predicts cleavage sites and a signal peptide. The server uses artificial neural networks and hidden Markov models.

View the [version history](#) of this server. All the previous versions are available.

New paper about using SignalP and other protein subcellular localisation predictors:
Locating proteins in the cell using TargetP, SignalP, and reLoc
Olaf Emanuelsson, Søren Brunak, Gunnar von Heijne, Henrik Nielsen
Nature Protocols 2, 953-971 (2007).

Access the paper and supplementary information [here](#).

SUBMISSION

Paste a single sequence or several sequences in [FASTA](#) format into the:

```
>P46644|AAT3_ARATH Aspartate aminotransferase, chloroplast -  
Arabidopsis thaliana (Mouse-ear cress).  
MKTTHFSSSSSSDRRIGALLRHLNSGSSDSDLMLSSLYASPTSGGTGGVYFSLVQAPEDPILGVTVAYMK
```

Submit a file in [FASTA](#) format directly from your local disk:

浏览...

Organism group

- Eukaryotes
- Gram-negative bacteria
- Gram-positive bacteria

Method

- Neural networks
- Hidden Markov models
- Both

Output format

- Standard
- Full
- Short (no graphics!)

Truncation

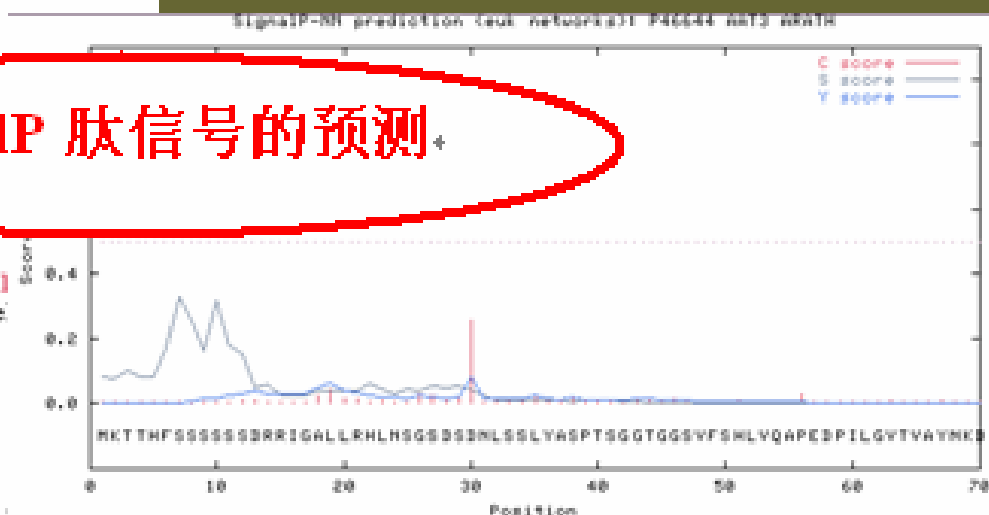
Truncate each sequence to max.

We recommend that only the N-term
Enter 0 (zero) to disable truncat

Submit

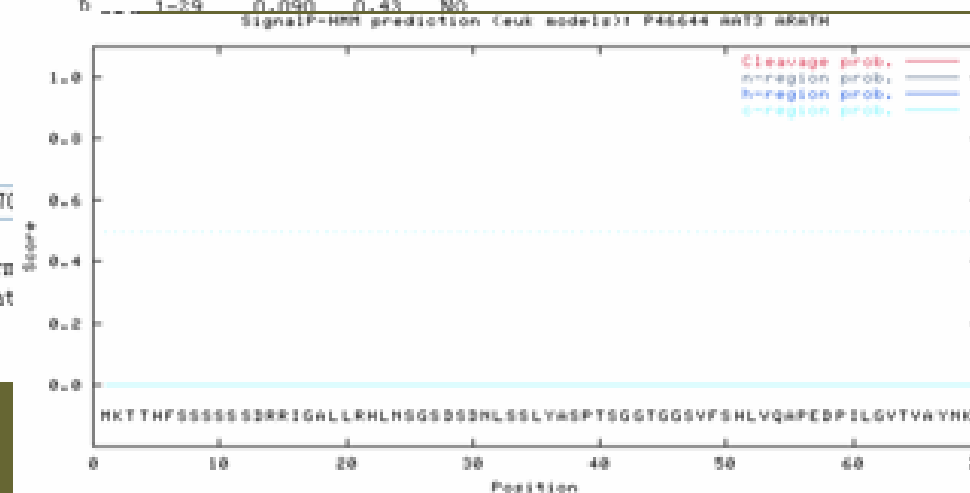
Clear fields

SignalP 肽信号的预测



data

```
>P46644_AAT3_ARATH length = 70  
# Measure Position Value Cutoff signal peptide?  
max. C 30 0.259 0.32 NO  
max. Y 30 0.086 0.33 NO  
max. S 7 0.331 0.67 NO  
mean S 1-29 0.095 0.48 NO  
n 1-29 0.090 0.43 NO
```



data

```
>P46644_AAT3_ARATH  
Prediction: 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 proteins based on the presence of a chloroplast transit peptide (cTP), mitochondrial targeting signal (mTP), and a presequence.

SUBMISSION

TargetP 亚细胞水平的预测

Paste a single sequence or several sequences in [FASTA](#) format into the field below:

>P46644|AAT3_ARATH Aspartate aminotransferase, chloroplast - Arabidopsis thaliana (Mouse-ear cress).
MKTTTFSSSSSSDERIGALLRHLNSGSDSDMLSSLYASPTSGGTGGSVFSHLVQAPEDPI

Submit a file in [FASTA](#) format directly from your local disk:

 浏览...

Organism group

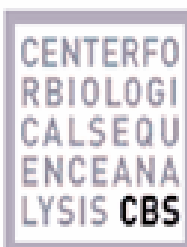
- Non-plant
 Plant

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: mTP:



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.
```

Name	Len	cTP	mTP	SP	other	Loc	RC	TPlen
P46644_AAT3_ARATH	449	0.545	0.310	0.036	0.149	C	4	64

NetNES 1.1 server predicts leucine-rich nuclear export signals (NES) in eukaryotic proteins using a combination of neural networks and hidden Markov models.

NetNES 亮氨酸富集式细胞核输出信号分析预测

Instructions Output format

SUBMISSION

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

```

MDSWYLLKRYELLCGLAGFTRLPQSLVLDMAVYLLDNYCPFERLLGQKATQADKDEE
ILSDIQPTLAVLTAQYQSLADNPELVISTEPSTPDPFFQVVALTSLSEELLARLQK
    
```

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

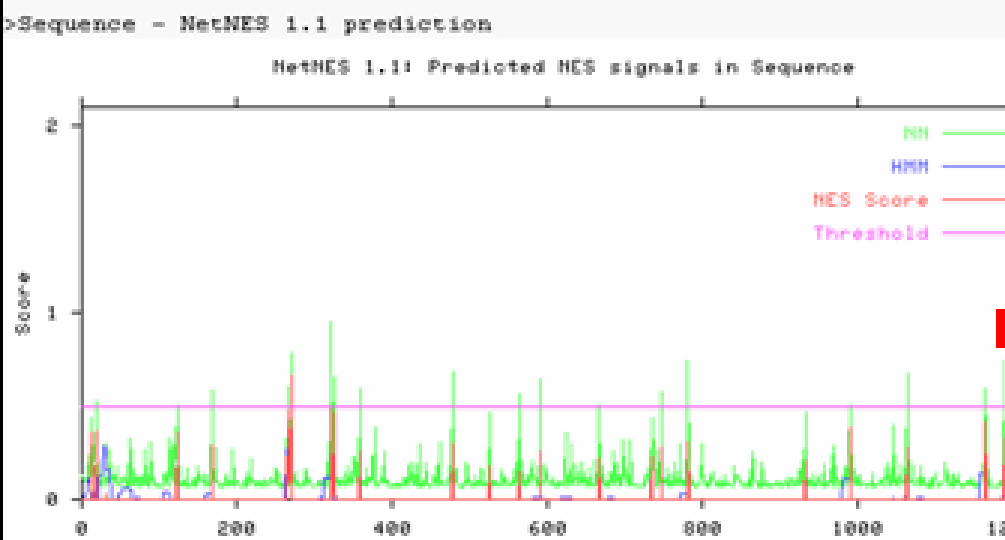
Generate graphics

NetNES 1.1 Server - prediction results

Technical University of Denmark

CENTER FOR BIOLOGICAL SEQUENCE ANALYSIS CBS

#Seq-Pos-Residue	Sequence Position			Predicted
	ANN	HMM	NES	
Sequence-1-M	0.124	0.000	0.000	-
Sequence-2-M	0.101	0.024	0.000	-
Sequence-3-R	0.100	0.024	0.000	-
Sequence-4-E	0.099	0.024	0.000	-
Sequence-5-M	0.108	0.024	0.000	-
Sequence-6-V	0.129	0.053	0.000	-
Sequence-7-L	0.148	0.104	0.000	-
Sequence-8-L	0.160	0.105	0.000	-
Sequence-9-M	0.239	0.107	0.000	-
Sequence-10-S	0.116	0.085	0.000	-
Sequence-11-V	0.255	0.085	0.000	-
Sequence-12-L	0.433	0.084	0.354	-
Sequence-13-L	0.378	0.079	0.000	-
Sequence-14-C	0.209	0.005	0.000	-
Sequence-15-G	0.133	0.005	0.000	-
Sequence-16-L	0.317	0.005	0.000	-
Sequence-17-A	0.205	0.000	0.000	-
Sequence-18-G	0.529	0.000	0.366	-
Sequence-270-R	0.078	0.267	0.103	-
Sequence-271-I	0.785	0.267	0.662	YES
Sequence-272-G	0.075	0.002	0.000	-
Sequence-273-E	0.113	0.002	0.000	-
Sequence-274-S	0.127	0.002	0.000	-
Sequence-275-D	0.086	0.002	0.000	-
Sequence-276-F	0.150	0.002	0.000	-
Sequence-277-F	0.074	0.002	0.000	-
Sequence-278-F	0.095	0.002	0.000	-
Sequence-279-T	0.078	0.000	0.000	-
Sequence-280-V	0.072	0.000	0.000	-
Sequence-281-P	0.074	0.000	0.000	-
Sequence-282-V	0.084	0.000	0.000	-
Sequence-283-S	0.075	0.000	0.000	-
Sequence-284-R	0.071	0.000	0.000	-



■ 兼有信号预测和跨膜分析功能

EMBL-EBI Search is currently closed. Please use the EBI search engines to find the data you need.

EMBL-EBI All Databases Enter Text Here Go Reset Advanced Search Give us feedback

Databases Tools EBI Groups Training Industry About Us Help Site Index

EBI > Tools > Protein Functional Analysis

Radar

RADAR stands for Rapid Automatic Detection of Repeats in Amino acid Residues. RADAR identifies repeats in protein sequences and structural units. Radar is uses an automatic algorithm to identify repeats, it identifies short composition biased as well as gapped repeats. Radar identifies architectures involving many different types of repeats in your query sequence.

Enter or Paste a Protein Sequence in any format

```

MESYKEIKIEDHPHYTLDMNFVAVTNSHSFTGLISDYTGVEGNNIMHTSSBLGFRELLGFQDLKCSSSA
SFFELPKEENSCPAVCVSEEVKPTAGESQNLISIVANQNVFNTPSTPNCSSISSEINEGHTNTTHED
AEAGEVLDHCGOQHTINTKQOLKAKKTVSQRKQREPRFAFRKSEVDFLEDGTRMRKYGQKAVRNSPFP
SNYYRCTSATCNVKKRVERCFSDPSIVVTTTEGKHTLSPMNTIHPPSCYPIITVVPASPGAFFLPRQ
FNINCSFMNLTSSLAMDQLDMAAFVAQGRRFCTSEMLGDEGLLQDLKPSSTLIKENTR
  
```

Upload a file:

Radar 重复序列信息的检测

Radar Results

Radar Results

Radar output: [radar-20080111-07410935375062.output](#)

No. of Repeats	Total Score	Length	Diagonal	BU-From	BU-To	Level
2	47.57	14	18	48	64	1
48-307	63 (21.24/24.41)	SSSLGFRELLgfQDLK				
	320 (26.33/13.97)	SEMLGDEGLL..QDLK				



ProtParam tool

ProtParam 蛋白质基础信息的统计

Protein name:
Molecular weight:
Theoretical pI:
Number of amino acids:
Molecular weight:
Theoretical pI:

Number of amino acids: 449
Molecular weight: 48954.3
Theoretical pI: 9.34

CSV format

Please note that you may only fill out **one** of the following fields at a time.

Enter a Swiss-Prot/TrEMBL accession number (AC) (for example P05130) or a sequence identifier (ID) (for example KPC1_DROME)

Or you can paste your own sequence in the box below:

```
QWQLEIRFMKSRSGLEKPPFTTSLAQTGQFASGSLDTHAKPISREPTVADGGECFLVAGSTLAKDGL  
T  
GERVQALSIVYKSLAPVAGRYESQLKLVIRPPTSSPPISGASIVAVTLRQKMLFWEWLE  
L  
KALADRIISRRQLFEALATGTPGQVSHIIRQLIGPFTFTGLMPAQVSPETKETHYET  
S  
DGRISRAGLSSSTVPELADATRAVYTKKY
```

RESET Compute parameters

composition:

Ala (A)	37	8.2%
Arg (R)	24	5.3%
Asn (N)	15	3.3%
Asp (D)	20	4.5%
Cys (C)	4	0.9%
Gln (Q)	15	3.3%
Glu (E)	18	4.0%
Gly (G)	39	8.7%
His (H)	13	2.9%
Ile (I)	28	6.2%
Leu (L)	45	10.0%
Lys (K)	25	5.6%
Met (M)	12	2.7%
Phe (F)	15	3.3%
Pro (P)	22	4.9%
Ser (S)	39	8.7%
Thr (T)	29	6.5%
Trp (W)	4	0.9%

Total number of negatively charged residues (Asp + Glu): 38
Total number of positively charged residues (Arg + Lys): 49

Atomic composition:

Carbon	C	2183
Hydrogen	H	3482
Nitrogen	N	606
Oxygen	O	639
Sulfur	S	16

Formula: C₂₁₈₃H₃₄₈₂N₆₀₆O₆₃₉S₁₆

- 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。

MENU

SwissModel First Approach Mode

Modeling requests:

- First Approach mode**
- Alignment interface
- Project (optimise) mode
- Oligomer modeling
- GPCR mode

Interactive tools

- DeepView - Swiss-PdbViewer, a tool for viewing and manipulating protein structures and models.
- Lookup the ExPDB template codes accessible to SWISS-MODEL.
- Search the template sequences accessible to SWISS-MODEL.
- Examples using SWISS-MODEL and the Swiss-PdbViewer.

HELP

Options:

- Frequently Asked Questions.
- Visualising 3D models.
- Reliability of models.
- How SWISS-MODEL works.
- How ProModII works.

Please fill these fields:

Your Email address: (MUST be correct!)

Your Name:

Request title: Will be added to the results header.

Provide a sequence or a SWISS-PROT AC code:

此处输入待测序列
(单字母氨基酸序列)

NOTES: A SWISS-PROT AC code looks like this: P04406
Sequences can be provided in either RAW, SWISS-PROT, FASTA or GCG format.

点击进入简捷模式

操作流程简图

注：预测结果将以邮件的形式发送给每一个客户。

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。

输入电子邮件地址（必须正确）

此处输入待测序列
(单字母氨基酸序列)

选择比对灵敏度

Lower BLAST limit:

- 0.0000000001
- 0.00000001
- 0.000001
- 0.00001**
- 0.0001

In some cases you may wish to define a set of template structures to be used for a modelling attempt. As an example, modelling a serpin (Serine esterase inhibitor similar to the plasminogen activator inhibitor, antithrombin III, etc) will generally fail, since these proteins have two distinct structures in the database:

1. the activated form of all true serpins
2. the precursor form as found in the serpin-analogue Ovalbumin.

It is thus best to choose the correct template(s) you wish to base your

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