Protein Search

Using the Protein Search

TAIR's Protein search allow you to search for proteins with a variety of parameters. You can perform a simple search by name, restrict your search to proteins having specific physio-chemial properties and domains,as well as limiting your search to proteins encoded in specified regions of the genome.

Search by Name

You can search for proteins by the following names:

- Gene name: For sequenced genes, the locus name corresponds to the orf name determined by AGI orf naming convention. For genetic loci (e.g. genes identified by mutation but not yet associated with a sequence) the name corresponds to the accepted symbolic name.AGI orf names have the format AT(1-5 or C,M)gXXXXXX.Where the value in parenthesis here corresponds to the chromosome number or organelle genome.
- Product Name: The name of the protein (e.g. Agamous protein or Dihydroflavanol 4-reductase).
- Gene Description: Brief summaries of the gene which encodes the protein of interest.This option is useful for including aspects of protein function or localization that are not indicated in the gene or product name.
- GenPeptID: Use this if you know the unique GenBank identifier for the protein.

Search by Structural Class Type

This feature allows you to restrict your search to include only those proteins belonging to the specified structural class . You may select multiple options within each parameter by clicking on one selection and then clicking on additional ones while holding down either the CTRL key (PCs) or the Apple key (Mac).

Structural class assignment was performed from annotations of SCOP's superfamilies using HMM models against TIGR's 4.0 Release by Drs JulianGough and Martin Madera at SCOP database. More information can be found in the following papers.

Gough, J., Hughey ,R., Karplus ,K., and Chothia ,C.(2001). Assignment ofgenome sequences using a library of hidden Markov models that represent all proteins of known structure. J Mol Biol. 2001 Nov 2;313(4):903-19

Gough, J., Chothia, C.(2002).SUPERFAMILY:HMMs representing all proteins of known structure. SCOP sequence searches, alignments, and genome assignments. Nucl. Acids Res., 2002 Jan 1;30(1):268-72

Search by Gene List

Use this option to make bulk queries using AGI locus IDs.

Search by Physio-chemical Properties

You can limit your search to include only proteins having specific physical/chemical properties.

Length

The calculated length of the translated protein in amino acids. This does not include lengths after processing such as cleavage of signal peptides.

Calculated MW

The predicted molecular weight in kiloDaltons. Molecular weights were calculated using the ATH1.pep file provided by TIGR including all predicted, experimentally verified and hypothetical proteins using the Bioperl function get_mol_wt (found in the SeqStats object at www.bioperl.org).

Calculated PI

The predicted isoelectric point.The isoelectric point is the point at which, on an isoelectric focusing gel the pH at which a protein has a net charge of zero. The pI of a protein is determined by its amino acid composition-and the net contribution of positive and negative charges of the side chains.The calculated pIs were determined using the ATH1.pep file from TIGR and the iep program from EMBOSS www.hgmp.mrc.ac.uk/Software/EMBOSS/.Values are between pH 1-14.

Domains

Use this option if you want to restrict your search to include only proteins having a specified domain composition.The drop down menus allow you to select the number of occurances of each domain along with the syntax for identifying the domain. For example, to search for proteins that have one or more occurances of domain PS00027 do the following:

- Select the greater than symbol in the first column.
- Enter one in the adjacent input box.
- Select Prosite from the domain type drop down menu
- Choose exact match
- Enter the name of the domain (PS00027)

If searching for more than one domain, the search is treated as a logical AND. Therefore, inputing PROSITE domain PS00027 AND PFAM domain PF04618 will limit your search to proteins having BOTH domains.

Protein domains are conserved regions of amino acid /structural similarity in protein sequences. Domains generally represent functional units having some form of biological activity. Domains are useful in grouping proteins with little overall sequence similarity.This search allows you to specify both the type and number of domains and use either the domain name , or unique domain identifier. Leaving this option blank will include all proteins in your search.The domains were identified using the interproscan.pl program and INTERPRO databases supplied with the program. For more information see: INTERPRO database.