CD-Search Element

Finds conserved domains in protein sequences. In case conserved domains database is downloaded the search can be executed on local machine. The search can be submitted to the NCBI for remote execution.

Element type: cd-search

Parameters

Parameter	Description	Default value	Parameter in Workflow File	Type
Annotate as	Name of the result annotations marking found conserved domains.	CDD result	**result-name**	*string*
Database	Currently, CD-Search is offered with the following search databases: • CDD - this is a superset including NCBI-curated domains and data imported from Pfam, SMART, COG, PRK, and TIGRFAM. • Pfam - a mirror of a recent Pfam-A database of curated seed alignments. Pfam version numbers do change with incremental updates. As with SMART, families describing very short motifs or peptides may be missing from the mirror. An HMM-based search engine is offered on the Pfam site. • SMART - a mirror of a recent SMART set of domain alignments. Note that some SMART families may be missing from the mirror due to update delays or because they describe very short conserved peptides and/or motifs, which would be difficult to detect using the CD-Search service. You may want to try the HMM-based search service offered on the SMART site. Note also that some SMART domains are not mirrored in CD because they represent "superfamilies" encompassing several individual, but related, domains; the corresponding seed alignments may not be available from the source database in these cases. Note also that SMART version numbers do not change with incremental updates of the source database (and the mirrored CD-Search database). • TIGRFAM - a mirror of a recent TIGRFAM set of domain alignments. An HMM-based search engine is offered on the TIGRFAM site. • COG - a mirror of the current COG database of orthologous protein families focusing on prokaryotes. Seed alignments have been generated by an automated process. An alternative search engine, "Cognitor", which runs protein-BLAST against a database of COG-assigned sequences, is offered on the COG site. • KOG - a eukaryotic counterpart to the COG database. KOGs are not included in the CDD superset, but are searchable as a separate data set.	C D D Available values are: • CDD • Pfam • TIGR FAM • COG • KOG • Prk • SMA RT	**db-name**	*string*
Database directory	Specifies database directory for local search.		**db-path**	*string*
Local search	Perform the search on local machine or submit the search to NCBI for remote execution.	True	**local-search**	*boolean*
Expect value	Modifies the E-value threshold used for filtering results. False positive results should be very rare with the default setting of 0.01, results with E-values in the range of 1 and above should be considered putative false positives.		**e-val**	*numeric*

Input/Output Ports

The element has 1 *input port*:

Name in GUI: *Input sequence*

Name in Workflow File: in-sequence

Slots:

Slot In GUI	Slot in Workflow File	Type
Sequence	**sequence**	*sequence*

And 1 *output port*:

Name in GUI: *Annotations*

Name in Workflow File: out-annotations

Slots:

Slot In GUI	Slot in Workflow File	Type
Set of annotations	**annotations**	*annotation-table*