

Scop3D – Manual

Where to find Scop3D

Scop3D is now available through an intuitive web-interface: <https://iomics.ugent.be/scop3d>

Example: Kras.

Kras is part of the Ras-Raf-MEK-ERK pathway. This pathway was the first one to be elucidated and remains an important research topic because of its central role in cellular differentiation in development. Because of this central role, inappropriate and/or continuous activation of this pathway is very common in cancer. Hence, several SNP variants are available for Kras.

Conservation analysis

Step 1: sequence analysis

homo sapiens	p01116
rattus norvegicus	p08644
mus musculus	p32883
danio rerio	q6aza4
pan troglodytus	h2q5m0
macaca mulatta	f6sfg7
felis catus	m3xau6
gallus gallus	f5ant2
ovis aries	w5qfh7
oryctolagus cuniculus	g1t2m0
bos taurus	e1bmx0

Sequence variants can be presented in fasta-format but can originate from several resources such as UniProt (The UniProt Consortium 2014 – www.uniprot.org) or NCBI (www.ncbi.nlm.nih.gov/protein). User-specific sequences can be used as well.

Fasta-file of these sequences

```
>sp|P01116|RASK_HUMAN GTPase KRas OS=Homo sapiens OX=9606 GN=KRAS PE=1 SV=1
MTEYKLVVVGAGGVGKSALTIQLIQNHVFDEYDPTIEDSYRKQVVIDGETCLLDILDITAG
QEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHHYREQIKRVKDSEDPMPVLVGNKCDL
PSRTVDTKQAQDLARSYGIPFIETSAKTRQQRVEDAFYTLVREIRQYRLKKISKEEKTGPGC
VKIKKCIIM
>sp|P08644|RASK_RAT GTPase KRas OS=Rattus norvegicus OX=10116 GN=Kras PE=1
SV=3
MTEYKLVVVGAGGVGKSALTIQLIQNHVFDEYDPTIEDSYRKQVVIDGETCLLDILDITAG
QEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHHYREQIKRVKDSEDPMPVLVGNKCDL
PSRTVDTKQAQDLARSYGIPFIETSAKTRQQRVEDAFYTLVREIRQYRLKKISKEEKTGPGC
VKIKKCVIM
>sp|P32883|RASK_MOUSE GTPase KRas OS=Mus musculus OX=10090 GN=Kras PE=1 SV=1
MTEYKLVVVGAGGVGKSALTIQLIQNHVFDEYDPTIEDSYRKQVVIDGETCLLDILDITAG
QEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHHYREQIKRVKDSEDPMPVLVGNKCDL
PSRTVDTKQAQDLARSYGIPFIETSAKTRQQRVEDAFYTLVREIRQYRLKKISKEEKTGPGC
VKIKKCVIM
>tr|Q6AZA4|Q6AZA4_DANRE K-ras OS=Danio rerio OX=7955 GN=kras PE=1 SV=1
MTEYKLVVVGAGGVGKSALTIQLIQNHVFDEYDPTIEDSYRKQVVIDGETCLLDILDITAG
QEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHHYREQIKRVKDSEDPMPVLVGNKCDL
QSHNVDSKQAQDLARSYGIPFIETSAKTRQGVDDAFYTLVREIRKHKEKMSKEGKKKKKK
SKTKCALM
>tr|H2Q5M0|H2Q5M0_PANTR KRAS isoform 2 OS=Pan troglodytes OX=9598 GN=KRAS
PE=2 SV=1
MTEYKLVVVGAGGVGKSALTIQLIQNHVFDEYDPTIEDSYRKQVVIDGETCLLDILDITAG
QEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHHYREQIKRVKDSEDPMPVLVGNKCDL
PSRTVDTKQAQDLARSYGIPFIETSAKTRQGVDDAFYTLVREIRKHKEKMSKDGKKKKKK
SKTKCVIM
>tr|F6SFG7|F6SFG7_MACMU KRAS proto-oncogene, GTPase OS=Macaca mulatta OX=9544
GN=KRAS PE=4 SV=1
MTEYKLVVVGAGGVGKSALTIQLIQNHVFDEYDPTIEDSYRKQVVIDGETCLLDILDITAG
QEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHHYREQIKRVKDSEDPMPVLVGNKCDL
PSRTVDTKQAQDLARSYGIPFIETSAKTRQQRVEDAFYTLVREIRQYRLKKISKEEKTGPGC
VKIKKCIIM
>tr|M3XAU6|M3XAU6_FELCA KRAS proto-oncogene, GTPase OS=Felis catus OX=9685
GN=KRAS PE=4 SV=2
MTEYKLVVVGAGGVGKSALTIQLIQNHVFDEYDPTIEDSYRKQVVIDGETCLLDILDITAG
QEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHHYREQIKRVKDSEDPMPVLVGNKCDL
PSRTVDTKQAQDLARSYGIPFIETSAKTRQGVDDAFYTLVREIRKHKEKMSKDGKKKKKK
SKTKCIIIFIAALDISVKASHQVTLVRPNII
>tr|F5ANT2|F5ANT2_CHICK K-Ras protein OS=Gallus gallus OX=9031 GN=K-RAS PE=2
SV=1
MTEYKLVVVGAGGVGKSALTIQLIQNHVFDEYDPTIEDSYRKQVVIDGETCLLDILDITAG
QEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHHYREQIKRVKDSEDPMPVLVGNKCDL
PSRTVDTKQAQDLARSYGIPFIETSAKTRQGVDDAFYTLVREIRKHKEKMSKDGKKKKKKE
TKTKCIIM
>tr|W5QFH7|W5QFH7_SHEEP KRAS proto-oncogene, GTPase OS=Ovis aries OX=9940
GN=KRAS PE=4 SV=1
MTEYKLVVVGAGGVGKSALTIQLIQNHVFDEYDPTIEDSYRKQVVIDGETCLLDILDITAG
QEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHHYREQIKRVKDSEDPMPVLVGNKCDL
PSRTVDTKQAQDLARSYGIPFIETSAKTRQQRVEDAFYTLVREIRQYRLKKISKEEKTGPGC
VKIKKCIVM
>tr|G1T2M0|G1T2M0_RABIT KRAS proto-oncogene, GTPase OS=Oryctolagus cuniculus
OX=9986 GN=KRAS PE=4 SV=1
GLPSPSPPPARRPGSRTPLPRSPALAGRPPAASPGHAAALAPGAQAGAGESLLKMTEYKL
VVVGAGGVGKSALTIQLIQNHVFDEYDPTIEDSYRKQVVIDGETCLLDILDITAGQEEYSA
MRDQYMRTGEGFLCVFAINNTKSFEDIHHYREQIKRVKDSEDPMPVLVGNKCDLPSRTVD
TKQAQDLARSYGIPFIETSAKTRQQRVEDAFYTLVREIRQYRLKKISKEEKTGPGCVKIKK
NVM
```

```
>tr|E1BMX0|E1BMX0_BOVIN KRAS proto-oncogene, GTPase OS=Bos taurus OX=9913
GN=KRAS PE=4 SV=2
MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDITAG
QEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHHYREQIKRVKDSEDPMPVLVGNKCDL
PSRTVDTKQAQDLARSYGIPFIETSAKTRQGVDDAFYTLVREIRKHKEKMSKDGKKKKKK
SKTKCIIM
```

To perform a conservation analysis of a protein or several sequence variants of protein, you have to fill the following section:

Sequence analysis

> Conservation analysis

☒

FASTA file

Browse...

FASTA file which contains all sequence variants of interest.

☐

Uniprot ID

Scop3D will perform a blast against the Uniprot database.

Identity cutoff

% Select all the matching sequences above this limit.

☐

Sequence

Scop3D will perform a blast against the Uniprot database.

Identity cutoff

% Select all the matching sequences above this limit.

In the **first option 'FASTA-file'**, you can upload a fasta-file with sequence variants of the protein of interest. An example of such a fasta-file is given on page 2.

In the **second option 'Uniprot ID'**, you can provide the uniprot ID of the protein you are interested in. For example: P01116.

This **identity cutoff** is of importance for the calculation of the consensus sequence which is calculated with the aid of BioPython. It specifies how common a particular residue has to be at a position before it is added to the consensus sequence. The default of this threshold is 0.7 (70%) as defined by BioPython. The usage of a lower value is advised when there are a low number of sequence variants.

In the **third option 'sequence'**, you can provide the protein sequence of the protein you are interested in. The **identity cutoff** has the same functionality as in the second option.

For example:

```
MTEYKLVVVGAGGVGKSALTIQLIQNHVFVDEYDPTIEDSYRKQVVIDGETCLLDILDITAG
QEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHHYREQIKRVKDSQVPMVLVGNKCDL
PSRTVDTKQAQDLARSYGIPFIETSAKTRQRVEDAFYTLVREIRQYRLKKISKEEKTGPGC
VKIKKCIIM
```

Step 2: structure annotation

Structure annotation

☒	BLAST	Scop3D will retrieve homologous structures by performing a BLAST search of the consensus sequence against the PDB database. You will be able to choose one of related PDB structures from the results screen.	☐	PDB ID		PDB ID of the protein to be used. If you have already the PDB file you want to be used please upload it below.
☐	Uniprot ID		☐	PDB file	Browse...	PDB file to be used. If you dont have the PDB file but you know the ID please fill it in above. No file chosen

SUBMIT

Here again, there are different options to add structural information.

The **first option 'blast'** can be used when you have no prior structural information. A BLAST (Johnson et al. 2008) against the PDB (Berman et al. 2000) using Biopython (Cock et al. 2009) with the consensus sequence as query sequence to identify the closest homologous structure is performed in this option.

In the **second option 'Uniprot ID'**, the user can provide a uniprot accession of the protein of interest. Scop3D then goes to uniprot and retrieves all structures available for that uniprot accession. The user can subsequently select a structure from this list to continue the analysis. For example: p01116

The **third option 'PDB ID'** can be used if the user knows which PDB-file (protein structure) her or she wants to work with. In our example, PDB-entry 4TQA can be used.

In the **last option 'PDB file'**, the user is able to submit his or her own PDB-file for the analysis.

If all the information needed is given, you can press the submit button and scop3D starts its calculations.

After you have submitted your job, you can follow the status of your job:

Your job was successfully submitted. The job ID is **54e7681f4a8f01acf38173e197d57998**. You can bookmark this page. The results will be available for 1 month.

Job status

Below you can find the status of your job:

- > BLASTing the UniProt database
- ⌚ Downloading the sequences from UniProt
- ⌚ Parsing sequences
- ⌚ Performing multiple sequence alignment
- ⌚ Splitting sequences file into individual sequence files
- ⌚ Performing pairwise alignment
- ⌚ Calculating identity and similarity percentages
- ⌚ Finding consensus sequence
- ⌚ Calculating abundance matrix
- ⌚ Calculating entropy values
- ⌚ Calculating conservation values
- ⌚ Drawing weblogo
- ⌚ Getting the PDB
- ⌚ Parsing the PDB
- ⌚ Performing chain sequence alignments

When the BLAST search is performed to determine the best matching structure, the user can choose which structure he wishes to continue with. The five best hits are given.

Important to notice is that the information about which chain that is homologous to your protein variants is given within the BLAST results.

If the PDB-entry is composed of multiple chains, the user has the option to choose which chain has to be adjusted. The output given in the BLAST results can be used by the user to guide the chain selection.

SNP variant analysis

Step 1: sequence analysis

In this example, we focus on the SNP variant analysis of Kras from *homo sapiens*. To perform an SNP variant analysis, the following section has to be filled:

> SNP variants analysis

☐ Uniprot ID
Scop3D will query the Uniprot database for SNP variants.

☐ DNA sequence
Scop3D will perform a blast against the NCBI nucleotide collection database.

Identity cutoff
% Select all the matching sequences above this limit.

☐ FASTA file
FASTA file which contains DNA sequences of interest.

The user has again several input options.

In the **first option 'Uniprot ID'**, the user has to provide the uniprot accession of the protein of interest; for example P01116 (for human Kras). Scop3D will subsequently retrieve the SNP variants available in UniProt and ensembl for this protein.

In the **second option 'DNA sequence'**, a DNA sequence can be provided. This sequence is subsequently used in BLAST search against the NCBI nucleotide collection database to identify homologous sequences. The identity cutoff allows you to retrieve sequences with a certain percentage of sequence identity.

While the **third option 'FASTA file'** allows the user to upload a fasta-file with his or her own DNA sequences that need to be analyzed.

Step 2: structure annotation

In this step, the user can follow the same input strategy as described in the conservation analysis.

Further help

In case you encounter problems with the usage of Scop3D or the interpretation of your results, you can contact the developers by email:

scop3d.compomics@vib-ugent.be