Scop3P Documentation

Scop3P is available as a web-interface and can be accessed at https://iomics.ugent.be/scop3p

Usage of Scop3P

Search Scop3P with Swiss-Prot accession/ID, protein name, PDB ID, ProteomeXchange ID or by keywords. For example, Epithelial growth factor receptor (EGFR) protein can be searched by typing ‘EGFR’ or ‘EGFR_HUMAN’ or ‘P00533’.

Importance of EGFR phosphorylation in cell cycle

EGFR is a transmembrane protein identified as a driver of tumorigenesis in different cancer types like lung and breast cancer and glioblastoma. Upon activation by specific ligands, EGFR dimerizes and autophosphorylates tyrosine residues that controls several signaling cascades and pathways such as PI3K/AKT and RAS/mitogen-activated protein kinase (MAPK) (Sigismund et al, 2018). Mutations of EGFR were found to be associated with several cancer types (Sigismund et al, 2018).

After the search, upon hit the result page shows information as interactive plots and tables. An overview of the P-sites for the searched protein will be displayed as shown below:

Overview panel
A) All the P-sites mapped onto the one dimensional amino acid sequence of the protein of interest (as ball and stick). The balls are colored based on the number of times the P-site is seen in different phosphoproteomics experiments. An empty ball denotes that this particular site is only annotated in Swiss-Prot and not seen in any of the proteomics experiments used in the Scop3P database. You can zoom in/out by using the mouse or by zoom options (+/-) on the top right corner of the panel.

B) P-site information such as sequence position, modified residue and the source information (obtained from UniProtKB, PRIDE projects or both) is displayed here. For the P-sites obtained from UniProtKB, the evidence details such as if the site is identified by homologous sequence similarity (Similarity) or by experimental identification (Experimental) or combination of both (Combined) will be shown.

C) The structure containing the most number of P-sites will be displayed by default. All P-sites are colored blue.

Tip: Mouse over the P-sites (ball and stick) on amino acid sequence to obtain more details and have a quick preview on where it is present on the protein structure (as shown in figure below). The amino acid in structure is highlighted and colored brown. For example the P-site at position 991 is mapped onto protein structure 3GOP at position 991 and is colored brown. The blue message panel shows the P-site is seen in 1 project and identified 9 times as phosphorylated. The structure can be visualized in different representations by adjusting the Display settings as shown in red box. Clicking Show All P-sites will display all P-sites mapped to the structure and colored blue.
Biophysical predictions

All predicted biophysical features can be viewed in the interactive circular plot as shown below. Every ring is annotated with different structural information.

The first amino acid in sequence and its associated predictions are colored dark in all rings. **Inner ring** contains the amino acid sequence and position information and the P-sites are colored red. Early folding values, backbone dynamics, disordered propensity, secondary structures are given in **ring 4,3 and 2** respectively.

A cut-off of 0.163 is used to classify early folding from non-early folding residues (>=0.163→**early folding** and <0.163→**non-early folding**), backbone dynamics were classified as **flexible** → <0.7, **context dependent** → 0.7-0.8 and **rigid** → >0.8, secondary structures were assigned a three class labels from DSSP, **H**→ helix, **E**→ sheet and **C**→ coil. For disordered propensity a cut-off measure of 0.5 is used to distinguish **disordered** (<0.5) and **ordered state** (>=0.5) The circular graph is interactive which can be zoomed in/out by mouse over the region of interest (inner panel in the circle), and can be reset to default by clicking the **Reset** button on top right corner. Mouse over the rings and the respective amino acids (and positions on the rings) will display the associated values.
Phospho-peptide table

This table shows all available phospho-peptides identified by re-processing different phosphoproteome experiments for the particular protein of interest. The number of projects column shows the occurrence of the particular peptide in different projects. For example P-site at position 1045 is seen in two different projects (PXD006482 and PXD000680). By clicking the drop down menu on anywhere on the column you can view the project details and the frequency of the peptide in that project.

Structure table:

This table will list all available structures and the P-sites mapped onto the structure. By default the structures are ordered based on the number of P-sites they has and then by resolution. For example here 3GOP has 6 P-sites and it is shown first in the table, while 3POZ and 3W33 have equal number of P-sites but are ordered based on the best resolution first.

Please note that some P-sites may not have a structure available, or they may be part of segments missing in the protein structure. You will not find such P-sites in this table, which only contains the sites which are seen in the observed segments of the protein structure. For example 3GOP chain A only contains P-sites 654,669,671,971,974,992 (e.g: 991 is in missing segment) and 3POZ chain A only contains 991,993,995,998,1016 (e.g 654,669,671 are not in modelled segment).

The method columns gives the structure determination method (X-ray, NMR etc) obtained from PDB, the stochiometry - interfacing molecule/ligand and CSS scores are PISA predicted. PISA uses different features to calculate the most probable assembly of a particular structure (See Methods section to know how it is calculated and the article referenced in the paper) to determine the multimeric state of the protein, which may not be the same as you see in PDB. CSS score ranges from 0-1 with 0 being less significant for the given complex (see Methods).
By clicking the dropdown icon or anywhere on the particular column, you can view the Secondary structural informations like secondary structure assigned by DSSP, conservation scale from 0-9 (1-less conserved to 9-highly conserved) from Consurf-DB and solvent accesibility assigned by PISA considering the interacting molecule/ligand in the biological assembly.

For instance P-site 995 (red box in inner panel) is in interface regions (as the buried surface area (BSA) is >0.0) and others are well exposed (BSA = 0) even though the structure appears to be monomeric. This is because the complex formation with chain A and ligand SO4 with chain A at position 1102 makes this position 995 less accessible.

Note that sometimes the sequence positions of P-sites will not be the same in all structures. For example in 3POZ (in the above picture) all UniProt and PDB positions are same but for 3POY (picture below) sequence positions 499 and 516 corresponds to 475 and 491 respectively.
By clicking the view structure icon (🔧) you can visualize the P-sites mapped onto that particular structure (see picture below). The structure can be viewed in different representations (panel A and B: surface, backbone, cartoon, spacefill and licorice) and alternatively P-sites can be colored based on accessibility, mutations (if any) and secondary structures (panel B). Color codes for accessibility in P-sites are cyan (exposed/surface), red (buried/core) and orange (crystal interface regions).
Mutation Table

This table gives all the amino acid variations (deleterious/neutral polymorphism) associated with the particular protein and the disease information as obtained from Humsavar data from UniprotKB.

<table>
<thead>
<tr>
<th>Uniprot Position</th>
<th>Amino acid (Wild Type)</th>
<th>Amino acid (Variant)</th>
<th>Variant Type</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>R</td>
<td>Q</td>
<td>Polypeptide</td>
<td>-</td>
</tr>
<tr>
<td>285</td>
<td>P</td>
<td>R</td>
<td>Polypeptide</td>
<td>-</td>
</tr>
<tr>
<td>420</td>
<td>D</td>
<td>D</td>
<td>Disease</td>
<td>Inflammatory skin and bone disease, normal, 2 (Absence) [OMIM: 136930]</td>
</tr>
<tr>
<td>521</td>
<td>R</td>
<td>K</td>
<td>Polypeptide</td>
<td>-</td>
</tr>
<tr>
<td>634</td>
<td>Y</td>
<td>I</td>
<td>Polypeptide</td>
<td>-</td>
</tr>
<tr>
<td>759</td>
<td>E</td>
<td>K</td>
<td>Polypeptide</td>
<td>-</td>
</tr>
<tr>
<td>779</td>
<td>D</td>
<td>A</td>
<td>Polypeptide</td>
<td>-</td>
</tr>
<tr>
<td>724</td>
<td>D</td>
<td>B</td>
<td>Polypeptide</td>
<td>-</td>
</tr>
<tr>
<td>754</td>
<td>E</td>
<td>K</td>
<td>Polypeptide</td>
<td>-</td>
</tr>
<tr>
<td>707</td>
<td>L</td>
<td>F</td>
<td>Undefined</td>
<td>-</td>
</tr>
</tbody>
</table>

Uniprot position, the wild type amino acid in the corresponding protein sequence, variation observed at this position and disease/polymorphism information will be displayed.

If any variations are on P-sites this can be viewed by coloring on the protein structure from Structure table.

Questions or suggestions?

For any further questions, feedback or suggestions, you can post an issue on the scop3P github page or send an email to one of the following persons:

Pathmanaban Ramasamy: pathmanaban.ramasamy [at] ugent [dot] be

Lennart Martens: lennart.martens [at] vib-ugent [dot] be

Wim Vranken: wim.vranken [at] vub [dot] be

References