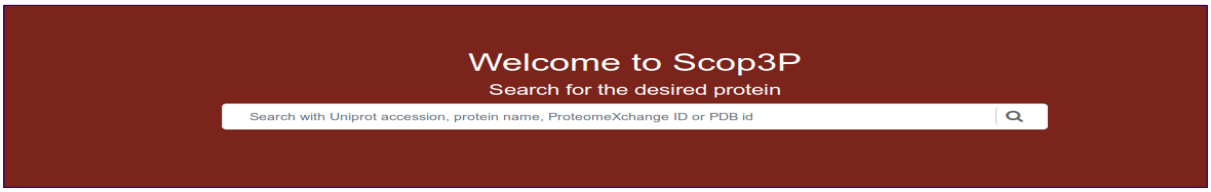
Scop3P Documentation

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**'.

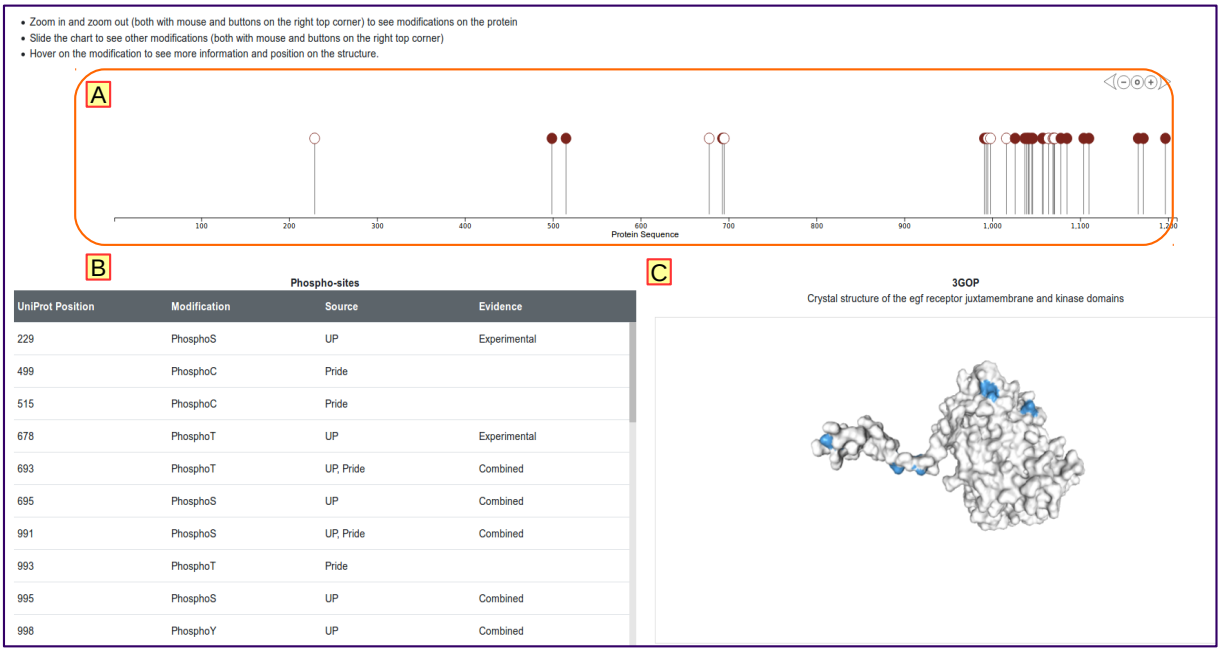


Importance of EGFR phopshorylation in cell cycle

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

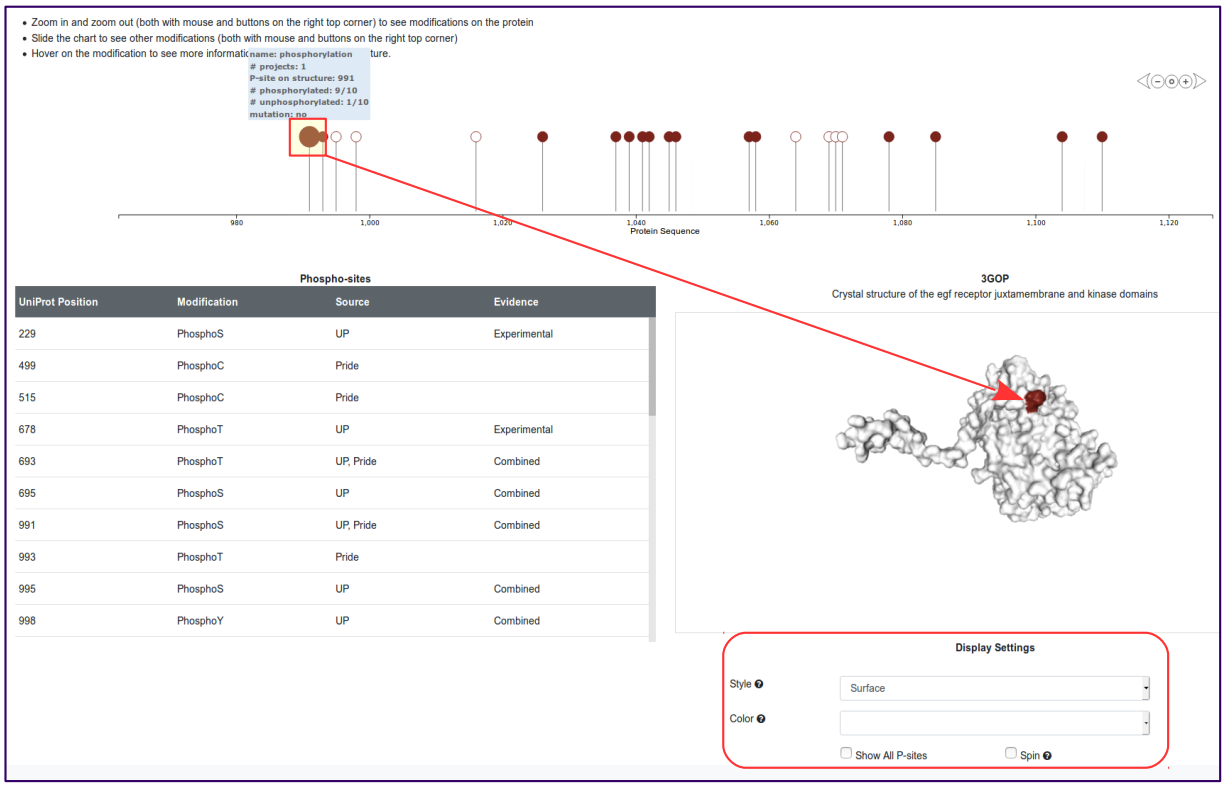


A) All the P-sites mapped onto the one dimensional amino acid sequence of the protein of interest (as ball and stick). The balls are colored based on the number of times the P-site is seen in different phosphoproteomics experiments. An empty ball denotes that this particular site is only annotated in 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.

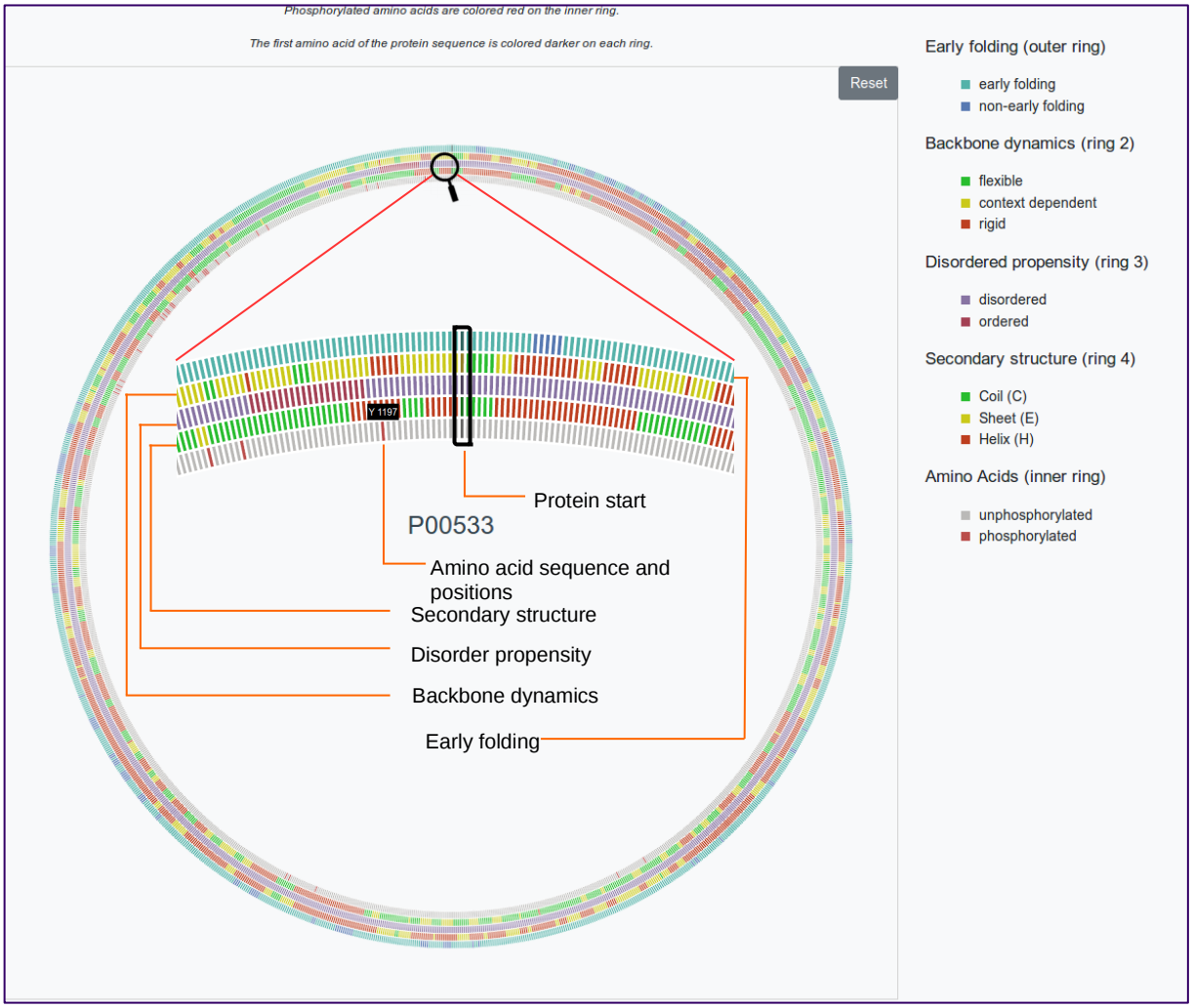
C) The structure containing the most number of P-sites will be displayed by default. All P-sites are colored blue.

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.

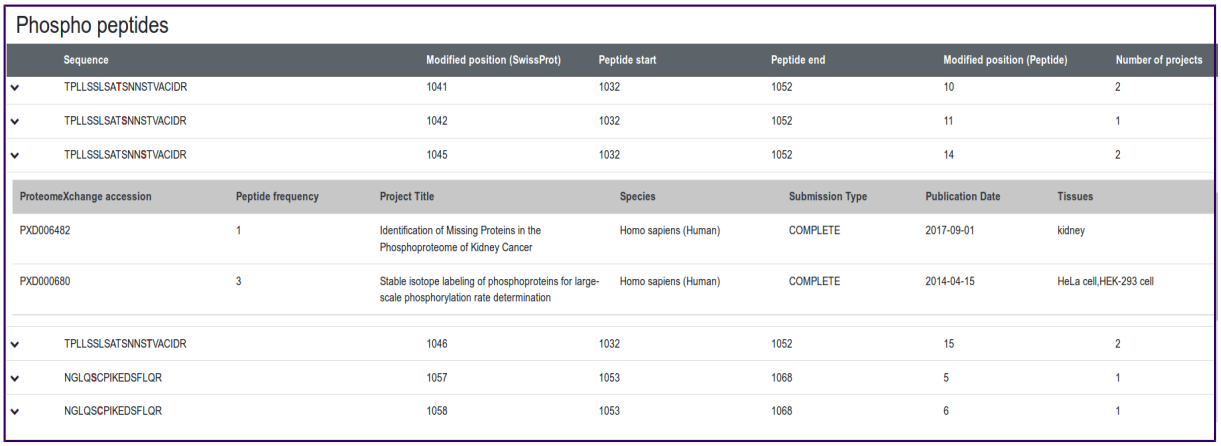


The first amino acid in sequence and its associated predictions are colored dark in all rings. **Inner ring** contains the amino acid sequence and position information and the P-sites are colored red. Early folding values, backbone dynamics, disordered propensity, secondary structures are given in **ring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



Phospho peptides

Sequence	Modified position (SwissProt)	Peptide start	Peptide end	Modified position (Peptide)	Number of projects
TPLLSSLSATSNNSTVACIDR	1041	1032	1052	10	2
TPLLSSLSATSNNSTVACIDR	1042	1032	1052	11	1
TPLLSSLSATSNNSTVACIDR	1045	1032	1052	14	2

ProteomeXchange accession	Peptide frequency	Project Title	Species	Submission Type	Publication Date	Tissues
PXD006482	1	Identification of Missing Proteins in the Phosphoproteome of Kidney Cancer	Homo sapiens (Human)	COMPLETE	2017-09-01	kidney
PXD000680	3	Stable isotope labeling of phosphoproteins for large-scale phosphorylation rate determination	Homo sapiens (Human)	COMPLETE	2014-04-15	HeLa cell,HEK-293 cell

Sequence	Modified position (SwissProt)	Peptide start	Peptide end	Modified position (Peptide)	Number of projects
TPLLSSLSATSNNSTVACIDR	1046	1032	1052	15	2
NGLQSCPIKEDSFLQR	1057	1053	1068	5	1
NGLQSCPIKEDSFLQR	1058	1053	1068	6	1

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).

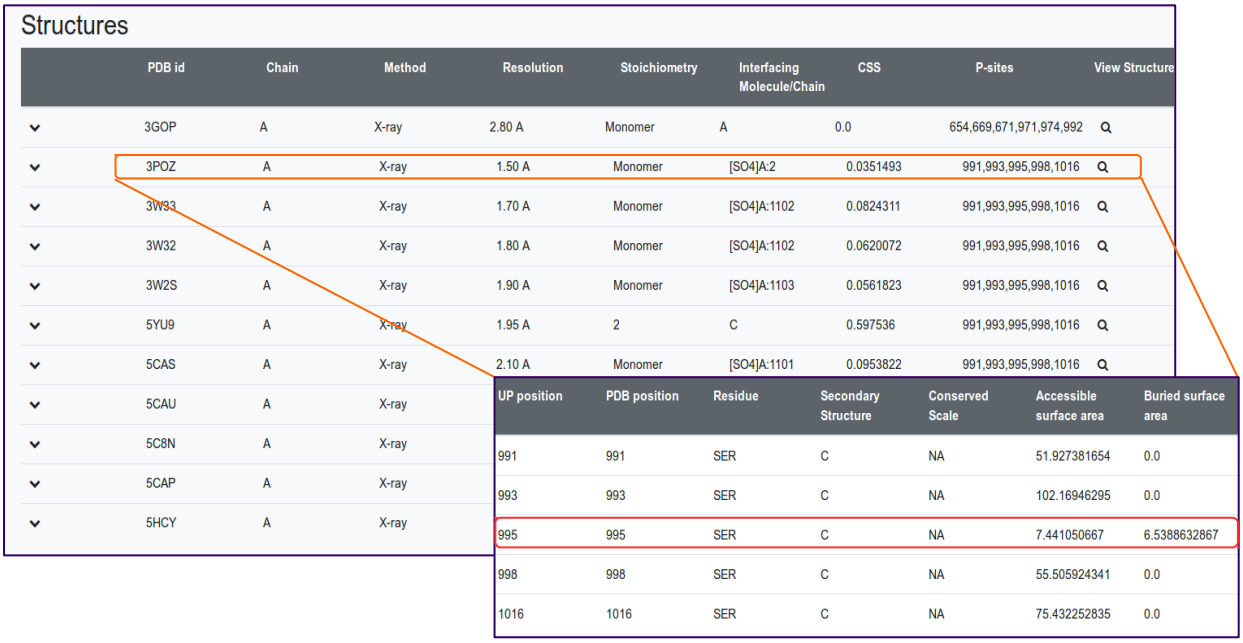
The **method** columns gives the structure determination method (X-ray, NMR etc) obtained from PDB, the stochiometry - interfacing molecule/ligand and **CSS** scores are **PISA** predicted. PISA uses different features to calculate the most probable assembly of a particular structure (See Methods section 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).

Structures

	PDB id	Chain	Method	Resolution	Stoichiometry	Interfacing Molecule/Chain	CSS	P-sites	View Structure
	3GOP	A	X-ray	2.80 A	Monomer	A	0.0	654,669,671,971,974,992	
	3POZ	A	X-ray	1.50 A	Monomer	[SO4]A:2	0.0351493	991,993,995,998,1016	
	3W33	A	X-ray	1.70 A	Monomer	[SO4]A:1102	0.0824311	991,993,995,998,1016	
	3W32	A	X-ray	1.80 A	Monomer	[SO4]A:1102	0.0620072	991,993,995,998,1016	
	3W2S	A	X-ray	1.90 A	Monomer	[SO4]A:1103	0.0561823	991,993,995,998,1016	
	5YU9	A	X-ray	1.95 A	2	C	0.597536	991,993,995,998,1016	
	5CAS	A	X-ray	2.10 A	Monomer	[SO4]A:1101	0.0953822	991,993,995,998,1016	
	5CAU	A	X-ray	2.25 A	Monomer	[SO4]A:1101	0.094314	991,993,995,998,1016	
	5C8N	A	X-ray	2.40 A	2	[SO4]A:1101	0.0873468	991,993,995,998,1016	
	5CAP	A	X-ray	2.40 A	2	[SO4]A:1101	0.372495	991,993,995,998,1016	
	5HCY	A	X-ray	2.46 A	2	[SO4]A:1101	0.0845778	991,993,995,998,1016	

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 posiiton 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 positons are same but for **3POY** (picture below) sequence positions 499 and 516 corresponds to 475 and 491 respectively.

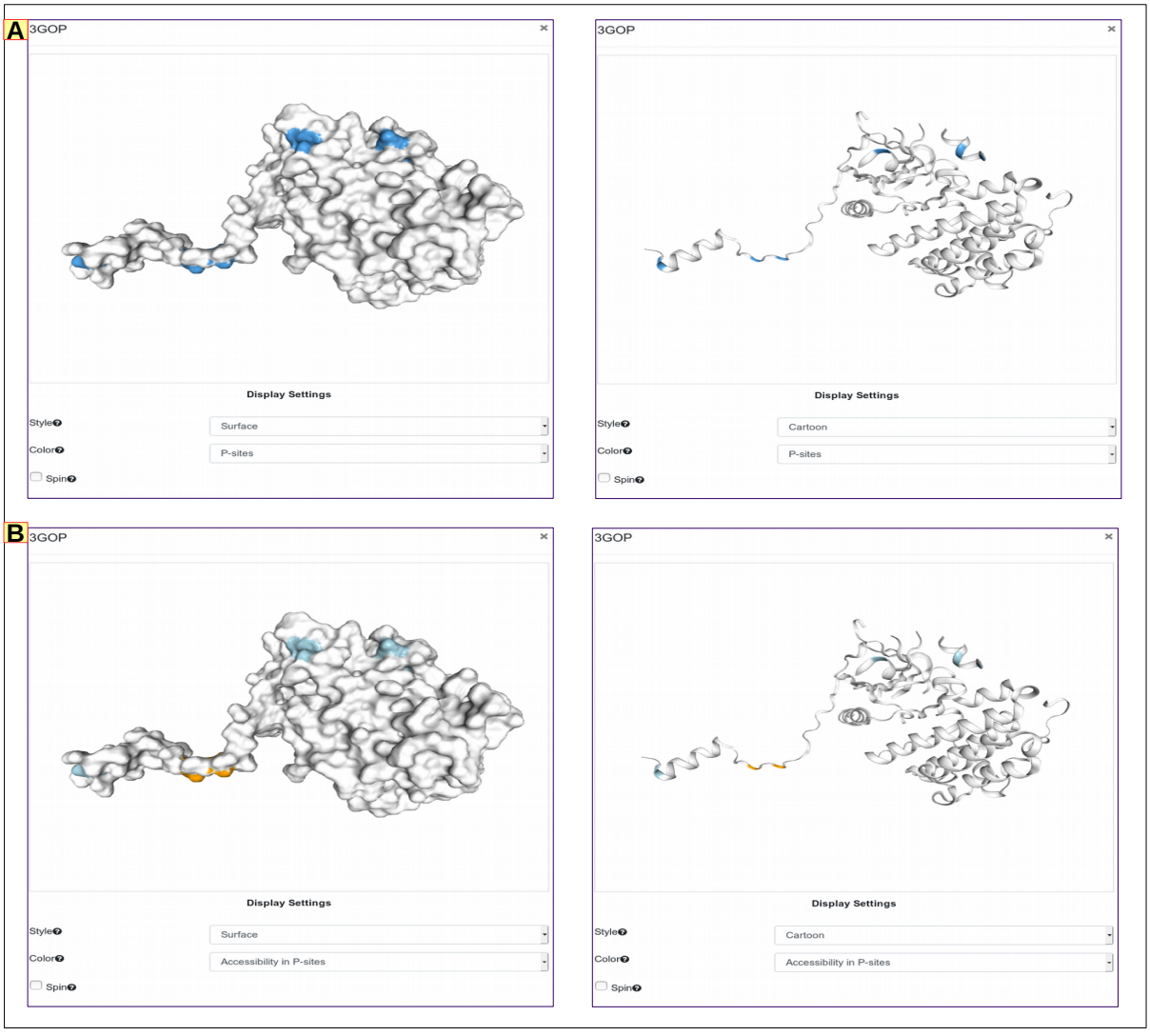
Structures

	PDB id	Chain	Method	Resolution	Stoichiometry	Interfacing Molecule/Chain	CSS	P-sites	View Structure
⌄	5WB8	A	X-ray	3.00 A	2	C	0.0	205,475,491	🔍
⌄	5X2K	A	X-ray	3.20 A	Monomer	[0UN]A:1101	0.1	995,998,1016	🔍
⌄	3NJP	A	X-ray	3.30 A	4	C	1.0	205,475,491	🔍
⌄	1IVO	A	X-ray	3.30 A	4	C	1.0	205,475,491	🔍
⌄	3P0Y	A	X-ray	1.80 A	2	H	0.0	475,491	🔍

UP position	PDB position	Residue	Secondary Structure	Conserved Scale	Accessible surface area	Buried surface area
499	475	CYS	E	9	6.8159984939	0.0
515	491	CYS	E	9	0.49371658916	0.0

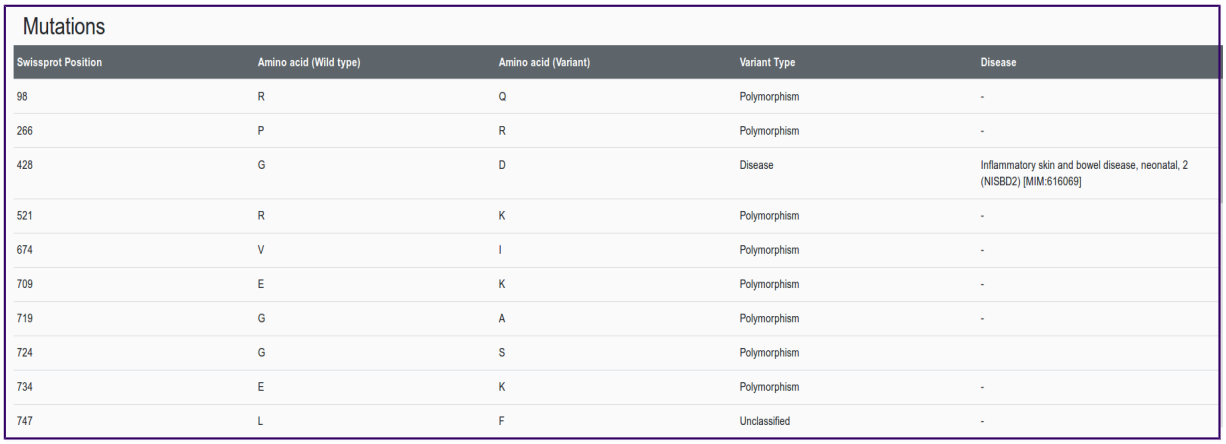
⌄	2RGP	A	X-ray	2.00 A	Monomer	[PO4]A:82	0.0689419	998,1016	🔍

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 UniprtoKB.



Mutations

Swissprot Position	Amino acid (Wild type)	Amino acid (Variant)	Variant Type	Disease
98	R	Q	Polymorphism	-
266	P	R	Polymorphism	-
428	G	D	Disease	Inflammatory skin and bowel disease, neonatal, 2 (NISBD2) [MIM:616069]
521	R	K	Polymorphism	-
674	V	I	Polymorphism	-
709	E	K	Polymorphism	-
719	G	A	Polymorphism	-
724	G	S	Polymorphism	
734	E	K	Polymorphism	-
747	L	F	Unclassified	-

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

Sigismund S, Avanzato D & Lanzetti L (2018) Emerging functions of the EGFR in cancer. *Mol. Oncol. 12: 3–20*