**Date: 03-03-22**

# WEBLEM 6
# Introduction to binding pocket prediction of protein w.r.t to PTM studies

Protein structures are complex and are sculpted with numerous surface pockets, internal cavities and cross channels. These topographic features provide structural basis and micro-environments for proteins to carry out their functions such as ligand binding, DNA interaction and enzymatic activity. Identification and quantification of these topographic features of proteins are therefore of fundamental importance for understanding the structure–function relationship of proteins, in engineering proteins for desired properties and in developing therapeutics against protein targets.

## CASTp:

- The CASTp server aims to provide comprehensive and detailed quantitative characterization of topographic features of proteins. Since its release 15 years ago, the CASTp server has ∼45,000 visits and fulfills ∼33,000 calculation requests annually. It has been proven to be a useful tool for a wide range of studies, including investigations of signaling receptors, discoveries of cancer therapeutics, understanding of mechanism of drug actions, studies of immune disorder diseases, analysis of protein–nanoparticle interactions, inference of protein functions and development of high-throughput computational tools.
- The CASTp server takes protein structures in the PDB format and a probe radius as input for topographic computation. Through the intuitive interface, users can either search for pre-computed results using a four-letter PDB ID, or submit their own protein structures to request customized computation. For pre-computed results, a default probe radius of 1.4 Å is used, which is the standard value for computing solvent accessible surface area. For customized computation request, users can specify any probe radius desired.
- The CASTp server identifies all surface pockets, interior cavities and cross channels in a protein structure and provides detailed delineation of all atoms participating in their formation. It also measures their exact volumes and areas, as well as sizes of the mouth openings if exist. These metrics are calculated analytically, using both the solvent accessible surface model (Richards' surface)and the molecular surface model (Connolly's surface). In addition, the CASTp server also provides imprints of topographic features. These results can be directly downloaded from CASTp server, which can be visualized using either the UCSF Chimera or our PyMOL plugin, CASTpyMOL.

## NetOGlyc – 4.0

- Glycosylation is the most abundant and diverse posttranslational modification of proteins. While several types of glycosylation can be predicted by the protein sequence context, and substantial knowledge of these glycoproteomes is available, our knowledge of the GalNAc-type O-glycosylation is highly limited. This type of glycosylation is unique in being regulated by 20 polypeptide GalNAc-transferases attaching the initiating GalNAc monosaccharides to Ser and Thr (and likely some Tyr) residues.
- The finding of unique subsets of O-glycoproteins in each cell line provides evidence that the O-glycoproteome is differentially regulated and dynamic. The greatly expanded view of the O-glycoproteome should facilitate the exploration of how site-specific O-glycosylation regulates protein function.

- The output conforms to the GFF version 2 format. For each input sequence the server prints a list of potential glycosylation sites, showing their positions in the sequence and the prediction confidence scores. Only the sites with scores higher than 0.5 are predicted as glycosylated and marked with the string "#POSITIVE" in the comment field.

## NetPhos - 3.1

- Protein phosphorylation at serine, threonine or tyrosine residues affects a multitude of cellular signaling processes. How is specificity in substrate recognition and phosphorylation by protein kinases achieved?
- In addition, serine and threonine residues in p300/CBP that can be modified by O-linked glycosylation with N-acetylglucosamine are identified. Glycosylation may prevent phosphorylation at these sites, a mechanism named yin-yang regulation.
- The results can be interpreted as:
    - **Sequence** - the sequence name;
    - # - the position of the residue in the sequence;
    - **x** - the residue in one-letter code;
    - **Context** - the sequence context of the residue, shown as a 9-residue subsequence centered on the residue;
    - **Score** - the prediction score (a value in the range [0.000-1.000]; the scores above **0.500** indicate positive predictions);
    - **Kinase** - the active kinase or the string "unsp" for non-specific prediction (as in NetPhos 2.0);
    - **Answer** - the string "**YES**" for positive predictions, else a dot.

## References:


- Tian, W., Chen, C., Lei, X., Zhao, J., & Liang, J. (2018). CASTp 3.0: Computed atlas of surface topography of proteins. Nucleic Acids Research, 46(W1). https://doi.org/10.1093/nar/gky473 Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology.
- Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT, Lavrsen K, Dabelsteen S, Pedersen NB, Marcos-Silva L, Gupta R, Bennett EP, Mandel U, Brunak S, Wandall HH, Levery SB, Clausen H. EMBO J, 32(10):1478-88, May 15, 2013. (doi: 10.1038/emboj.2013.79. Epub 2013 Apr 12)
- Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. Blom, N., Gammeltoft, S., and Brunak, S. Journal of Molecular Biology: 294(5): 1351-1362, 1999.

**Date: 03-03-22**

# WEBLEM 6A
# To predict binding pocket of protein Glutamine using Castp server.

## Introduction:

- The CASTp server aims to provide comprehensive and detailed quantitative characterization of topographic features of proteins. Since its release 15 years ago, the CASTp server has ~45,000 visits and fulfills ~33,000 calculation requests annually. It has been proven to be a useful tool for a wide range of studies.
- **Glutamine**
    - Glutamine is the most abundant and versatile amino acid in the body. In health and disease, the rate of glutamine consumption by immune cells is similar or greater than glucose. For instance, in vitro and in vivo studies have determined that glutamine is an essential nutrient for lymphocyte proliferation and cytokine production, macrophage phagocytic plus secretory activities, and neutrophil bacterial killing. Glutamine release to the circulation and availability is mainly controlled by key metabolic organs, such as the gut, liver, and skeletal muscles.

## Methodology:

- Take a PDB id from a protein structure of Glutamine
- Enter the PDB id on the webpage of CASTp
- Interpret the results

## Observation:

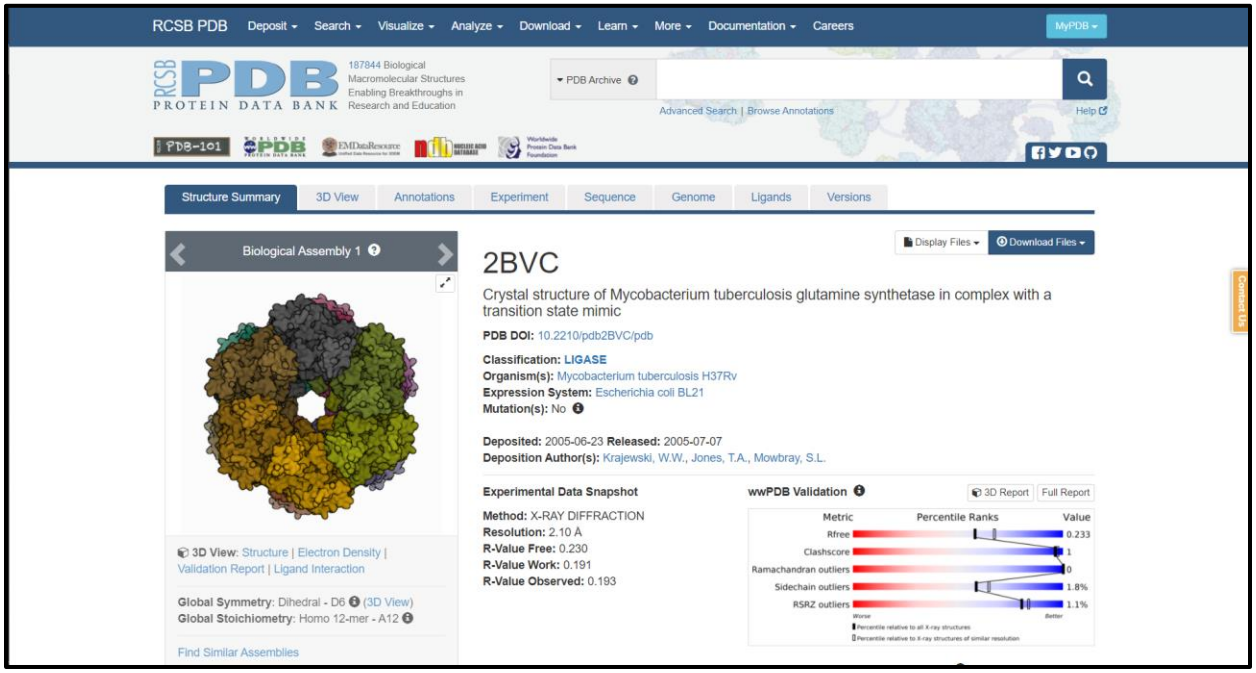



**Fig1. PDB page for query Glutamine**

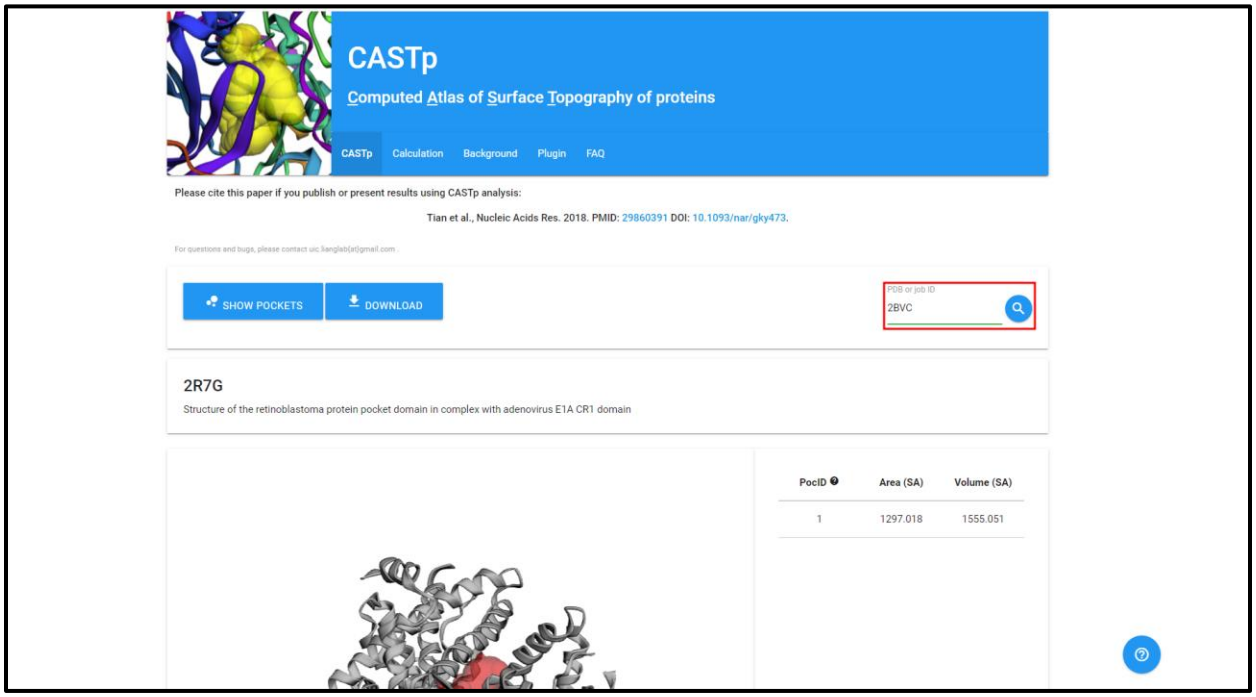



**Fig2. CASTp page with the PDB id of my query entered**

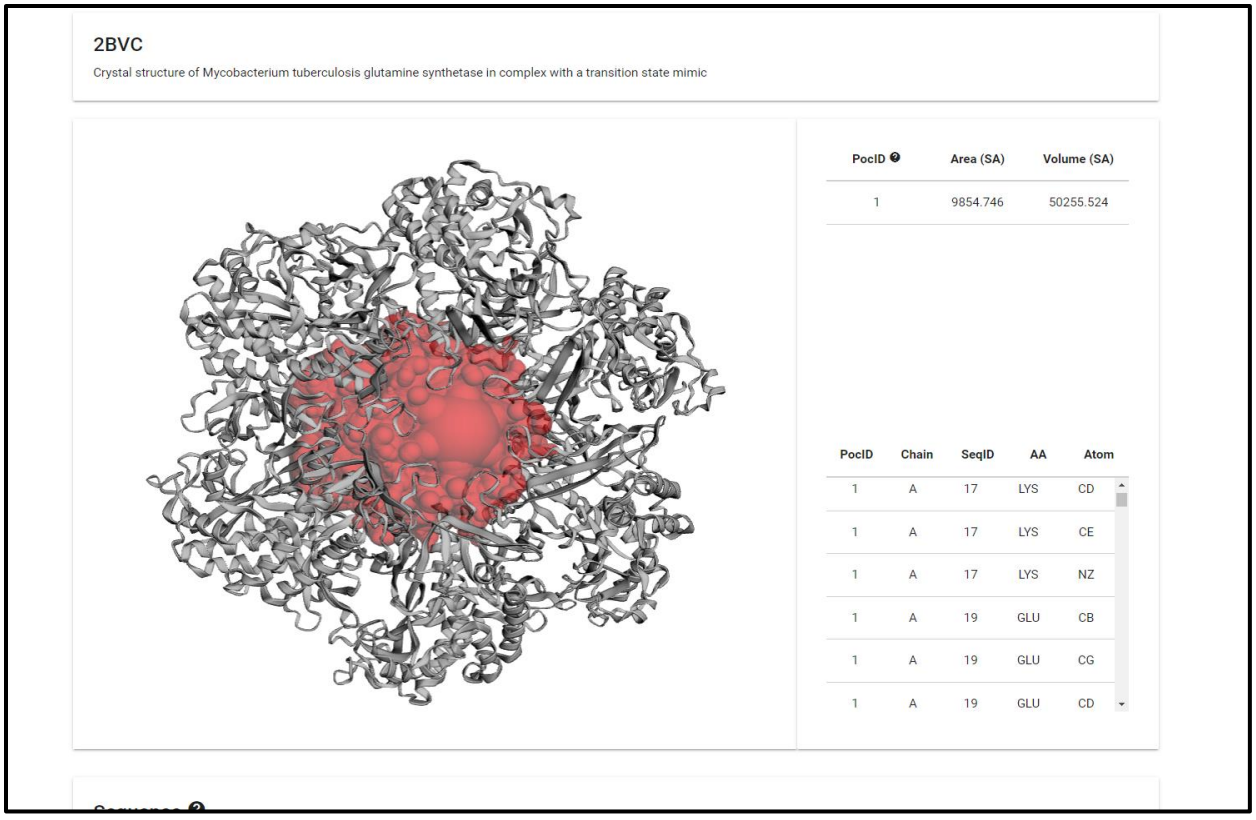



**Fig3. Result of my query on CASTp**

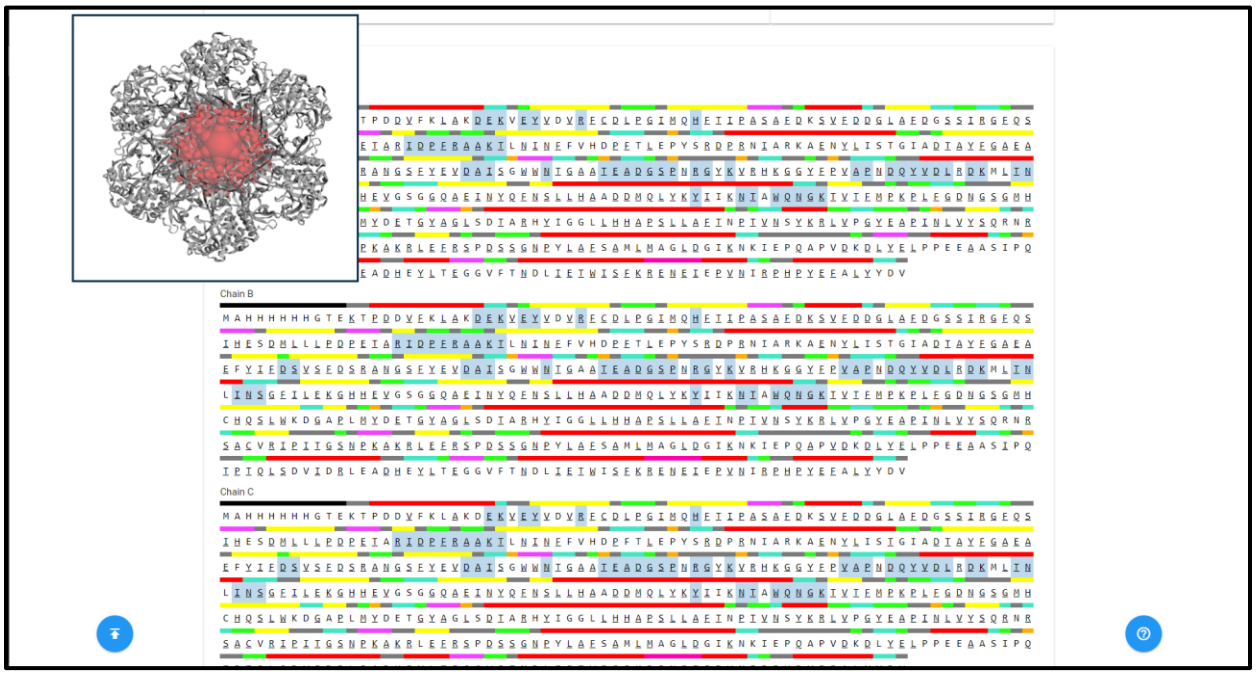
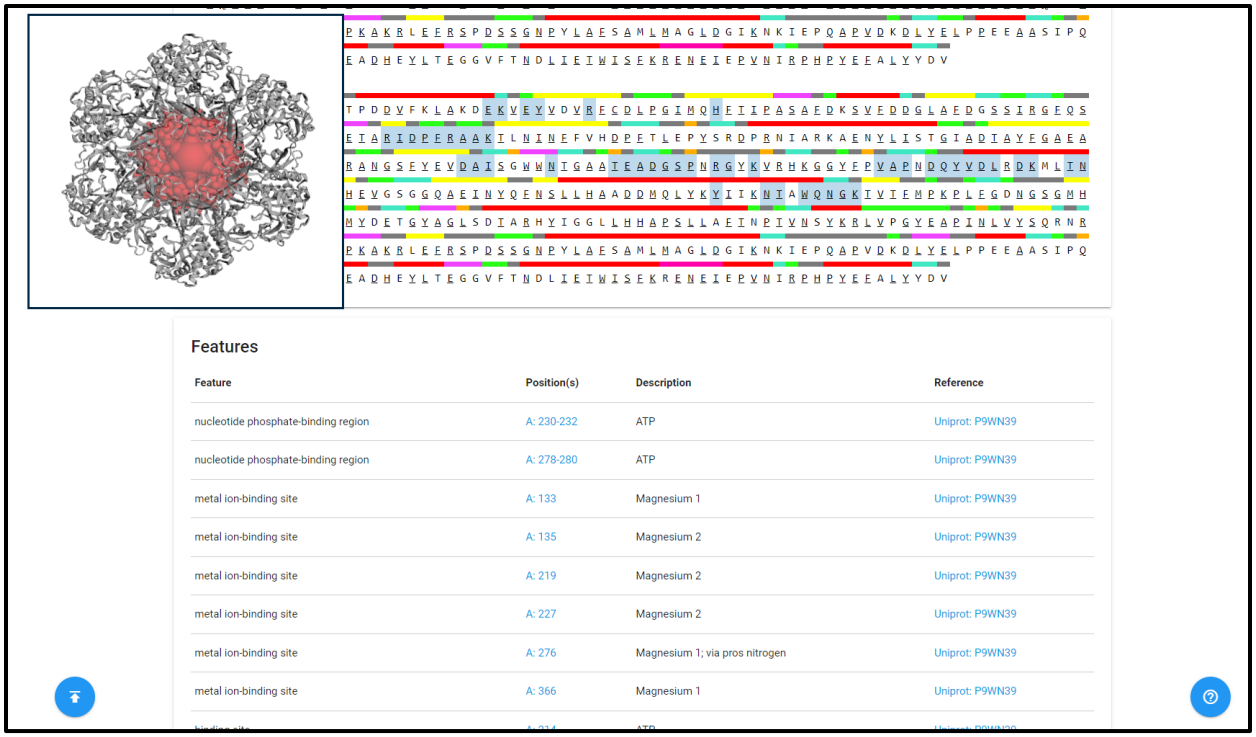


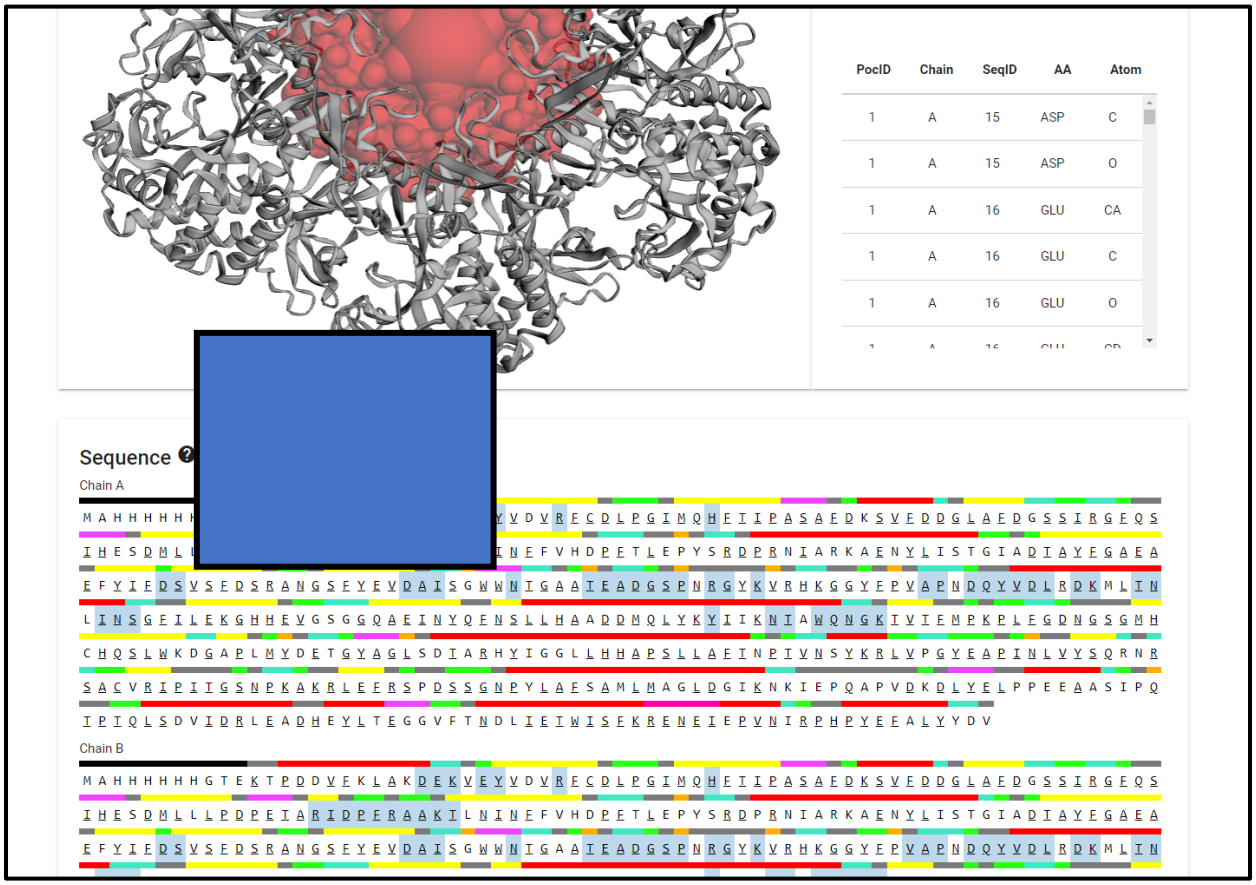

**Fig4. Sequence data of the query on CASTp**



Features

| Feature | Position(s) | Description | Reference |
|---|---|---|---|
| nucleotide phosphate-binding region | A: 230-232 | ATP | Uniprot: P9WN39 |
| nucleotide phosphate-binding region | A: 278-280 | ATP | Uniprot: P9WN39 |
| metal ion-binding site | A: 133 | Magnesium 1 | Uniprot: P9WN39 |
| metal ion-binding site | A: 135 | Magnesium 2 | Uniprot: P9WN39 |
| metal ion-binding site | A: 219 | Magnesium 2 | Uniprot: P9WN39 |
| metal ion-binding site | A: 227 | Magnesium 2 | Uniprot: P9WN39 |
| metal ion-binding site | A: 276 | Magnesium 1; via pros nitrogen | Uniprot: P9WN39 |
| metal ion-binding site | A: 366 | Magnesium 1 | Uniprot: P9WN39 |

**Fig5. Features data of my query on CASTp**



| PocID | Chain | SeqID | AA | Atom |
|---|---|---|---|---|
| 1 | A | 15 | ASP | C |
| 1 | A | 15 | ASP | O |
| 1 | A | 16 | GLU | CA |
| 1 | A | 16 | GLU | C |
| 1 | A | 16 | GLU | O |

Sequence

Chain A

Chain B

**Fig6. Legend to refer to interpret data**

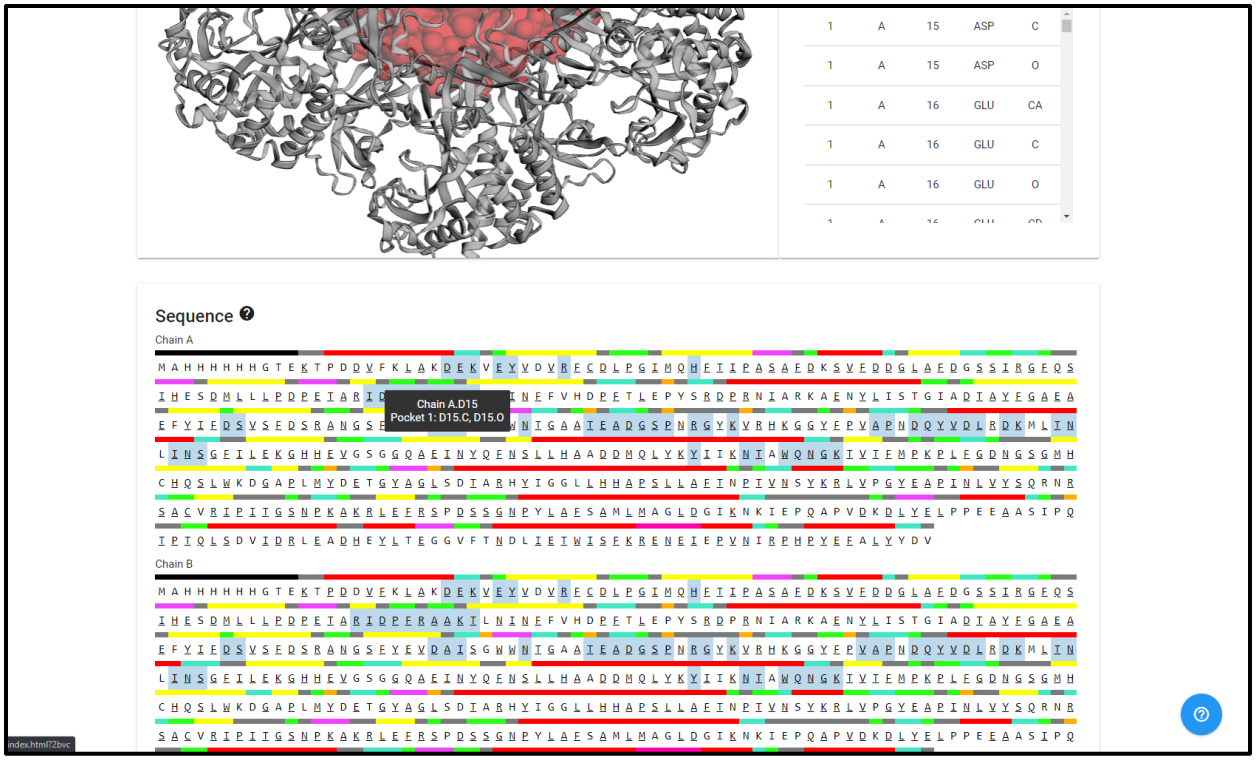



**Fig7. Pocket info for Chain A.D15, Pocket 1: D15.C, D15.O**


**Fig8. Position in 3D space of nucleotide phosphate binding region A:230-232**

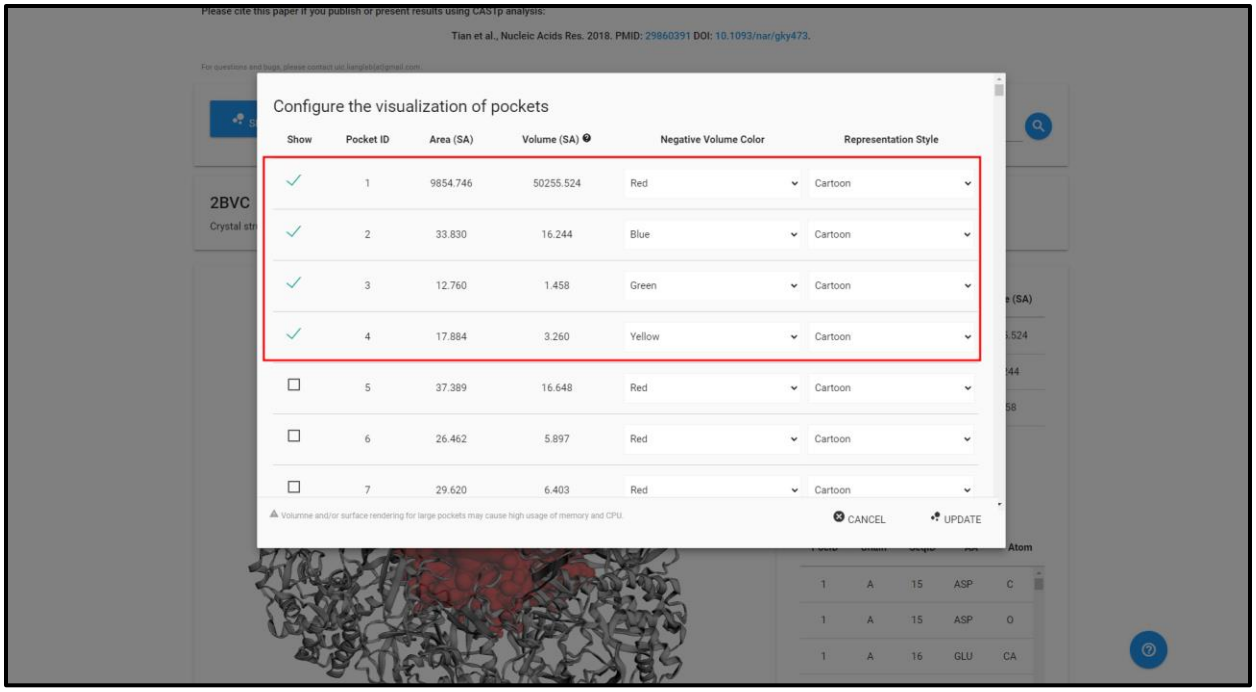



**Fig9. Customizing the visualization of pocket in CASTp**

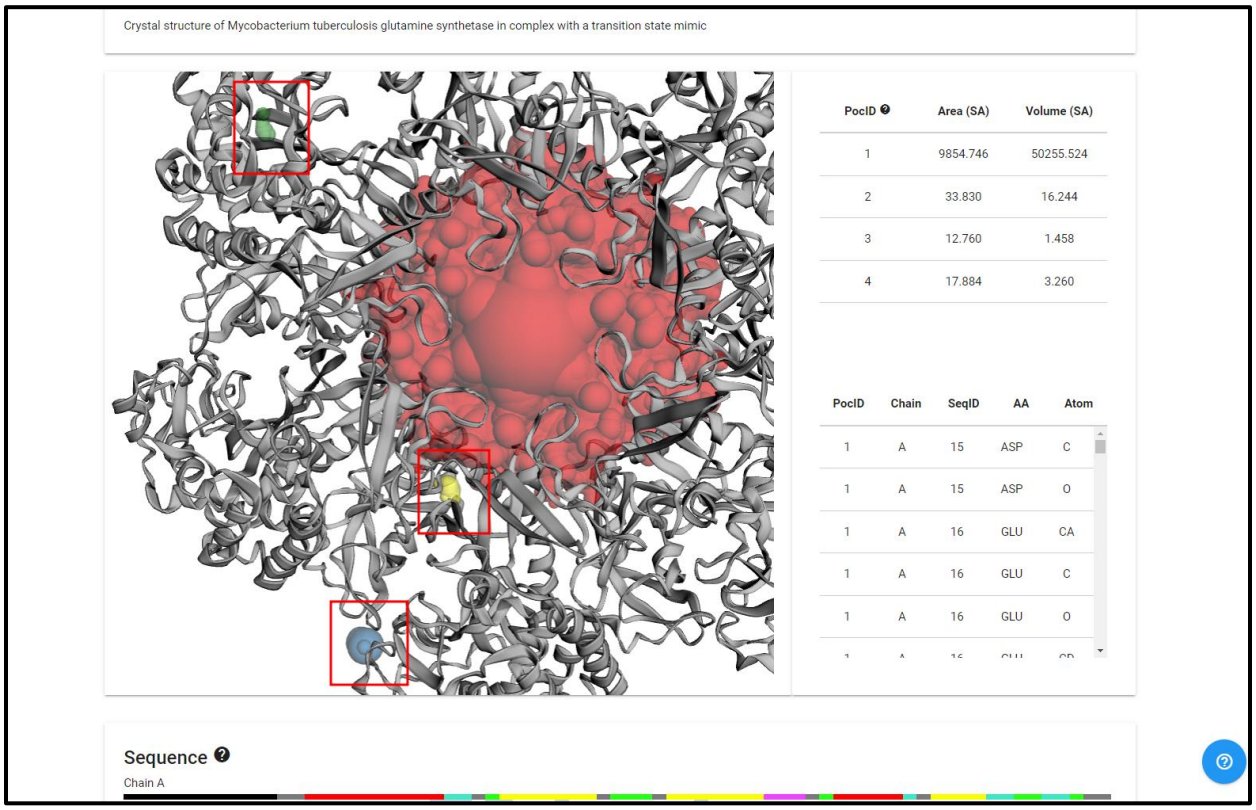



**Fig10. Customized view of pockets for query in CASTp**

**Result:**

1. The Results are categorically divided:
    a. **Structure:** Here we can customize the visualization of the protein structure.
    b. **Sequence:** Here we can see the amino acid sequences of the given protein structure.
    c. **Features:** Here we see the features of the provided protein structure
2. In the given structure the Area is 9854.746 SA and the volume is 50255.524 SA

## Observation:

- Castp is a great tool to find out the specific sites on the protein in 3D space.
- The colors of the site can be changed according to us and be seen clearly and boldly.

## References:


- Bank, R. C. S. B. P. D. (n.d.). *2BVC: Crystal structure of mycobacterium tuberculosis glutamine synthetase in complex with a transition state mimic*. RCSB PDB. Retrieved March 3, 2022, from https://www.rcsb.org/structure/2BVC
- CASTp 3.0: Computed atlas of surface topography of proteins. (n.d.). Retrieved March 3, 2022, from http://sts.bioe.uic.edu/castp/index.html?2bvc
- Cruzat, V., Macedo Rogero, M., Noel Keane, K., Curi, R., & Newsholme, P. (2018). Glutamine: Metabolism and immune function, supplementation and clinical translation. *Nutrients*, *10*(11), 1564. https://doi.org/10.3390/nu10111564

**Date: 03-03-22**

# WEBLEM 6B
# To predict binding pocket for Glycosylation sites in (query name) using NetOGlyc 4.0 Server

## Introduction:

- Glycosylation is the most abundant and diverse posttranslational modification of proteins.
- The output conforms to the GFF version 2 format. For each input sequence the server prints a list of potential glycosylation sites, showing their positions in the sequence and the prediction confidence scores. Only the sites with scores higher than 0.5 are predicted as glycosylated and marked with the string "#POSITIVE" in the comment field.
- **Glutamine**
    - Glutamine is the most abundant and versatile amino acid in the body. In health and disease, the rate of glutamine consumption by immune cells is similar or greater than glucose. For instance, in vitro and in vivo studies have determined that glutamine is an essential nutrient for lymphocyte proliferation and cytokine production, macrophage phagocytic plus secretory activities, and neutrophil bacterial killing. Glutamine release to the circulation and availability is mainly controlled by key metabolic organs, such as the gut, liver, and skeletal muscles.

## Methodology:

- Take a FASTA sequence from uniport
- Enter the sequence into the submission box
- Interpret the result according to the output tab

## Observation:

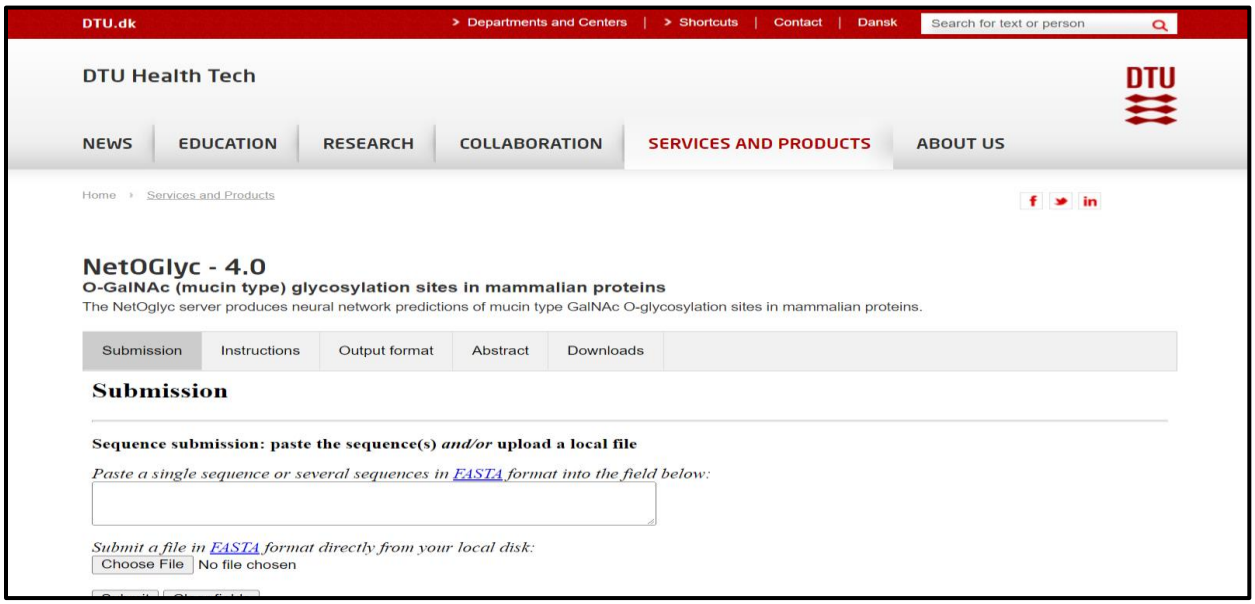
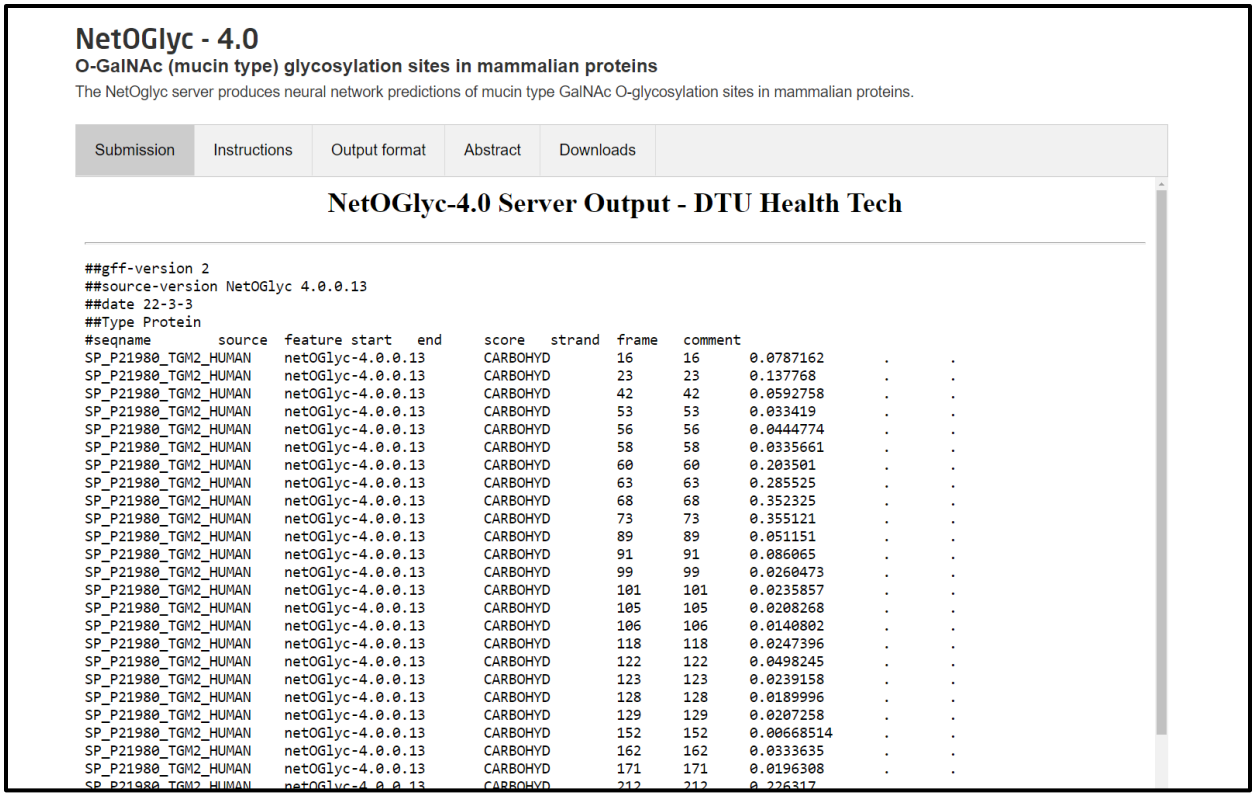



**Fig1. Homepage of NetOGlyc – 1.0**

```
>sp|P21980|TGM2_HUMAN Protein-glutamine gamma-glutamyltransferase 2 OS=Homo sapiens
OX=9606 GN=TGM2 PE=1 SV=2
MAEELVLERCDLELETNGRDHHTADLCREKLVVRRGQPFWLTLHFEGRNYEASVDSLTFS
VVTGPAPSQEAGTKARFPLRDAVEEGDWTATVVDQQDCTLSLQLTTPANAPIGLYRLSLE
ASTGYQGSSFVLGHFILLFNAWCPADAVYLDSEEERQEYVLTQQGFIYQGSAKFIKNIPW
NFGQFEDGILDICLILLDVNPKFLKNAGRDCSRRSSPVYVGRVVSGMVNCNDDQGVLLGR
WDNNYGDGVSPMSWIGSVDILRRWKNHGCQRVKYGQCWVFAAVACTVLRCLGIPTRVVTN
YNSAHDQNSNLLIEYFRNEFGEIQGDKSEMIWNFHCWVESWMTRPDLQPGYEGWQALDPT
PQEKSEGTYCCGPVPVRAIKEGDLSTKYDAPFVFAEVNADVVDWIQQDDGSVHKSINRSL
IVGLKISTKSVGRDEREDITHTYKYPEGSSEEREAFTRANHLNKLAEKEETGMAMRIRVG
QSMNMGSDFDVFAHITNNTAEEYVCRLLLCARTVSYNGILGPECGTKYLLNLNLEPFSEK
SVPLCILYEKYRDCLTESNLIKVRALLVEPVINSYLLAERDLYLENPEIKIRILGEPKQK
RKLVAEVSLQNPLPVALEGCTFTVEGAGLTEEQKTVEIPDPVEAGEEVKVRMDLLPLHMG
LHKLVVNFESDKLKAVKGFRNVIIGPA
```

**Fig2. FASTA Sequence for query Glutamine**


**Fig3. Homepage of NetOGlyc with the FASTA sequence of query pasted in the search box**



NetOGlyc - 4.0

O-GalNAc (mucin type) glycosylation sites in mammalian proteins

The NetOglyc server produces neural network predictions of mucin type GalNAc O-glycosylation sites in mammalian proteins.

Submission | Instructions | Output format | Abstract | Downloads

NetOGlyc-4.0 Server Output - DTU Health Tech

```
##gff-version 2
##source-version NetOGlyc 4.0.0.13
##date 22-3-3
##Type Protein
#seqname        source  feature start   end     score   strand  frame   comment
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        16      16      0.0787162       .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        23      23      0.137768        .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        42      42      0.0592758       .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        53      53      0.033419        .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        56      56      0.0444774       .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        58      58      0.0335661       .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        60      60      0.203501        .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        63      63      0.285525        .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        68      68      0.352325        .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        73      73      0.355121        .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        89      89      0.051151        .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        91      91      0.086065        .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        99      99      0.0260473       .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        101     101     0.0235857       .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        105     105     0.0208268       .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        106     106     0.0140802       .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        118     118     0.0247396       .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        122     122     0.0498245       .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        123     123     0.0239158       .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        128     128     0.0189996       .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        129     129     0.0207258       .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        152     152     0.00668514      .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        162     162     0.0333635       .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        171     171     0.0196308       .       .
SP_P21980_TGM2_HUMAN    netOGlyc-4.0.0.13       CARBOHYD        212     212     0.226317
```

**Fig4. Result page of NetNGlyc showing the submission data for query**

NetOGlyc - 4.0

O-GalNAc (mucin type) glycosylation sites in mammalian proteins

The NetOglyc server produces neural network predictions of mucin type GalNAc O-glycosylation sites in mammalian proteins.

Submission | Instructions | Output format | Abstract | Downloads

## Output format

### DESCRIPTION

The output conforms to the GFF version 2 format. For each input sequence the server prints a list of potential glycosylation sites, showing their positions in the sequence and the prediction confidence scores [redacted] and marked with the string "#POSITIVE" in the comment field.

The example below shows the output for human granulocyte-macrophage colony-stimulating factor, The example below shows the output for human granulocyte-macrophage colony-stimulating factor, taken from the UniProt entry CSF2_HUMAN. Currently, 4 sites have been experimentally annotated for this protein, and NetOGlyc predicts that two of these are glycosylated. Additionally, it predicts an additional site is glycosylated at site 108. Occupancy of O-glycosylation sites can vary in-vivo depending on the cells that are expressing the protein. The interactions between sites of initial O-Glycosylation with subsequent sites of glycosylation are yet to be fully elucidated, while our capability to precisely predict the substrate specificity of individual GalNAc-Ts remains limited. The combination of these factors mean that although NetOGlyc will attempt to predict individual sites of glycosylation, a safe interpretation of a positive prediction is that the protein in that local region is more likely to carry O-GalNAc modifications.

### EXAMPLE OUTPUT

```
##gff-version 2
##source-version NetOGlyc 4.0.0.12
##date 13-7-15
##Type Protein
#seqname        source  feature start   end     score   strand  frame   comment
CSF2_HUMAN      netOGlyc-4.0.0.12       CARBOHYD        5       5       0.04656 .       .
```

**Fig5. Output format for NetOGlyc used to refer to interpret the results of our query**

## Result:

- After submitting the FASTA sequence for query Glutamine it gives 3 potential glycolysis sites.
- This interpretation can be made by looking at the scores higher than 0.5 and a comment that has #POSITIVE

## Conclusion:

- To conclude NetOGlyc is a great source to gather data about the potential Glycolysis sites quickly and accurately
- We can submit one or multiple sequences upto 50 sequences and 200,000 amino acids at once which makes the processing of sequences more efficient

## References:


- Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT, Lavrsen K, Dabelsteen S, Pedersen NB, Marcos-Silva L, Gupta R, Bennett EP, Mandel U, Brunak S, Wandall HH, Levery SB, Clausen H. EMBO J, 32(10):1478-88, May 15, 2013. (doi: 10.1038/emboj.2013.79. Epub 2013 Apr 12)
- Uniprot. (n.d.). Retrieved March 3, 2022, from https://www.uniprot.org/uniprot/P21980.fasta
- Cruzat, V., Macedo Rogero, M., Noel Keane, K., Curi, R., & Newsholme, P. (2018). Glutamine: Metabolism and immune function, supplementation and clinical translation. *Nutrients*, *10*(11), 1564. https://doi.org/10.3390/nu10111564

**Date: 03-03-22**

# WEBLEM 6B
# To predict binding pocket for Phosphorylation site in (query name) using NetPhos 3.1 server

## Introduction:

- Protein phosphorylation at serine, threonine or tyrosine residues affects a multitude of cellular signaling processes. How is specificity in substrate recognition and phosphorylation by protein kinases achieved? In addition, serine and threonine residues in p300/CBP that can be modified by O-linked glycosylation with N-acetylglucosamine are identified. Glycosylation may prevent phosphorylation at these sites, a mechanism named yin-yang regulation.
- **Glutamine**
    - Glutamine is the most abundant and versatile amino acid in the body. In health and disease, the rate of glutamine consumption by immune cells is similar or greater than glucose. For instance, in vitro and in vivo studies have determined that glutamine is an essential nutrient for lymphocyte proliferation and cytokine production, macrophage phagocytic plus secretory activities, and neutrophil bacterial killing. Glutamine release to the circulation and availability is mainly controlled by key metabolic organs, such as the gut, liver, and skeletal muscles.

## Methodology:

- Take FASTA sequence from uniprot
- Enter it in the submission box for NetPhos
- Interpret the result according to the output page of NetPhos

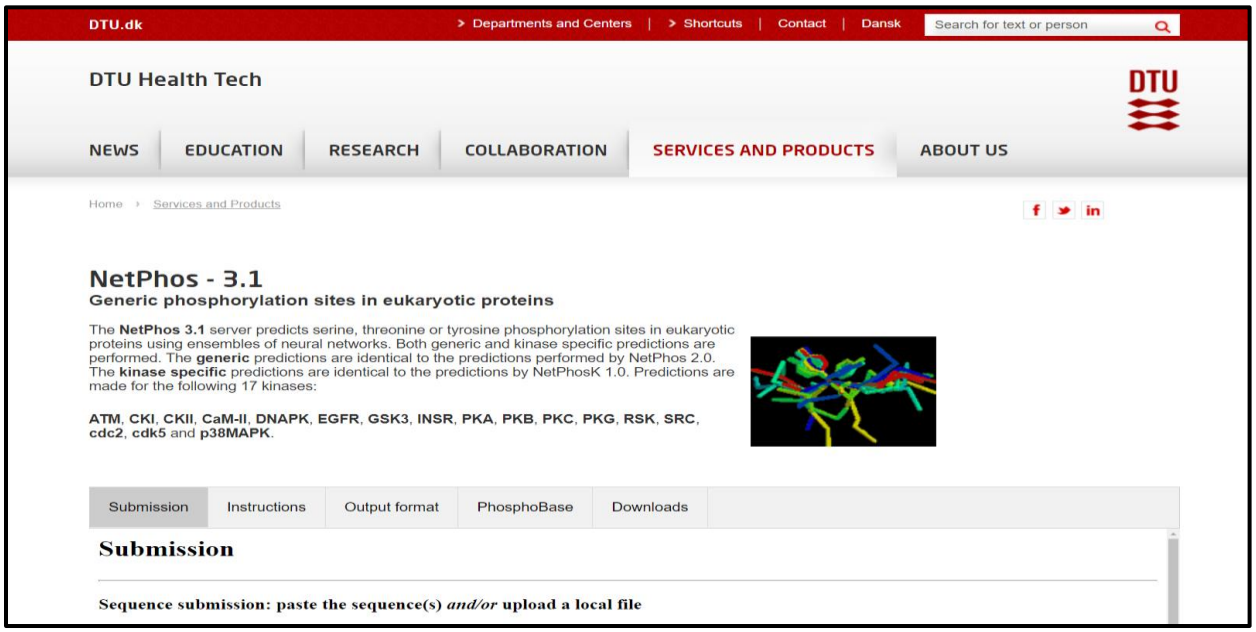



**Fig1. Homepage of NetPhos 3.1**

```
>sp|P21980|TGM2_HUMAN Protein-glutamine gamma-glutamyltransferase 2 OS=Homo sapiens
OX=9606 GN=TGM2 PE=1 SV=2
MAEELVLERCDLELETNGRDHHTADLCREKLVVRRGQPFWLTLHFEGRNYEASVDSLTFS
VVTGPAPSQEAGTKARFPLRDAVEEGDWTATVVDQQDCTLSLQLTTPANAPIGLYRLSLE
ASTGYQGSSFVLGHFILLFNAWCPADAVYLDSEEERQEYVLTQQGFIYQGSAKFIKNIPW
NFGQFEDGILDICLILLDVNPKFLKNAGRDCSRRSSPVYVGRVVSGMVNCNDDQGVLLGR
WDNNYGDGVSPMSWIGSVDILRRWKNHGCQRVKYGQCWVFAAVACTVLRCLGIPTRVVTN
YNSAHDQNSNLLIEYFRNEFGEIQGDKSEMIWNFHCWVESWMTRPDLQPGYEGWQALDPT
PQEKSEGTYCCGPVPVRAIKEGDLSTKYDAPFVFAEVNADVVDWIQQDDGSVHKSINRSL
IVGLKISTKSVGRDEREDITHTYKYPEGSSEEREAFTRANHLNKLAEKEETGMAMRIRVG
QSMNMGSDFDVFAHITNNTAEEYVCRLLLCARTVSYNGILGPECGTKYLLNLNLEPFSEK
SVPLCILYEKYRDCLTESNLIKVRALLVEPVINSYLLAERDLYLENPEIKIRILGEPKQK
RKLVAEVSLQNPLPVALEGCTFTVEGAGLTEEQKTVEIPDPVEAGEEVKVRMDLLPLHMG
LHKLVVNFESDKLKAVKGFRNVIIGPA
```

**Fig2. FASTA Sequence for Query Glutamine**

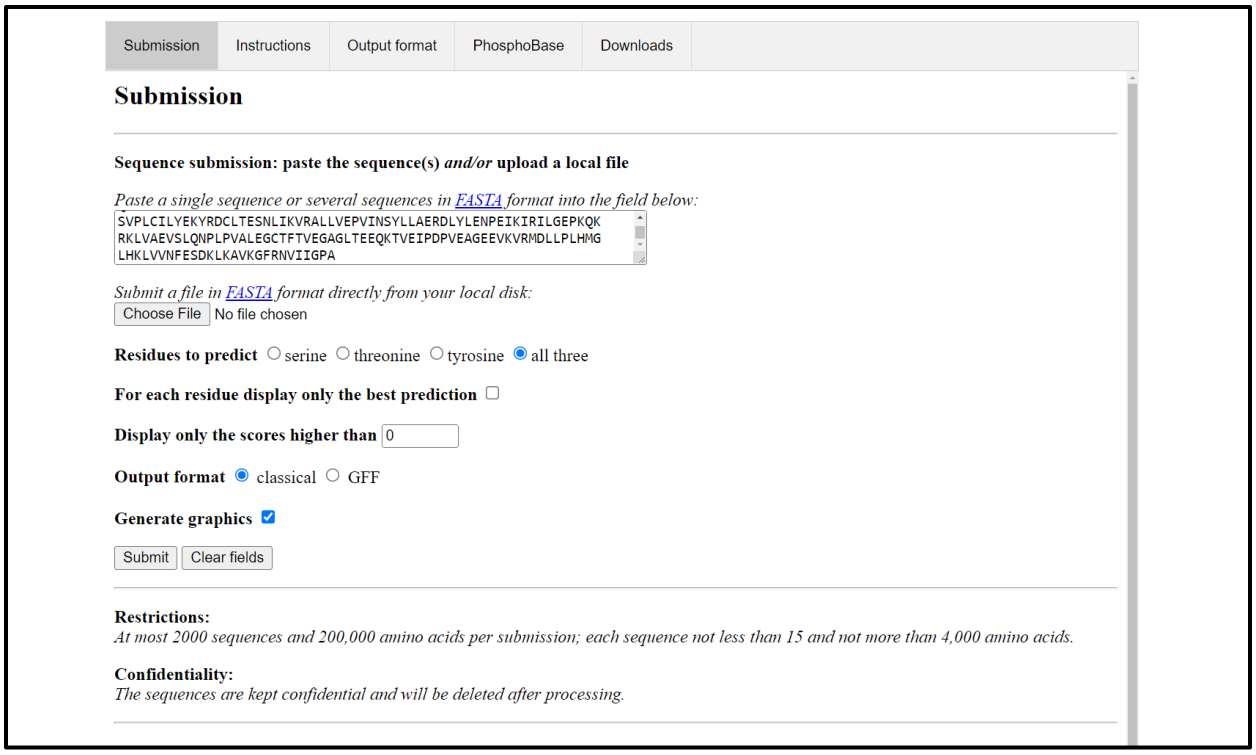



**Fig3. FASTA sequence pasted in the submission box on NetPhos 3.1**

Submission | Instructions | Output format | PhosphoBase | Downloads

**NetPhos-3.1 Server Output - DTU Health Tech**

```
>sp_P21980_TGM2_HUMAN   687 amino acids
#
# netphos-3.1b prediction results
#
# Sequence                 # x   Context     Score   Kinase    Answer
# ---------------------------------------------------------------------
# sp_P21980_TGM2_HUMAN    16 T   LELETNGRD   0.486   CKII        .
# sp_P21980_TGM2_HUMAN    16 T   LELETNGRD   0.461   GSK3        .
# sp_P21980_TGM2_HUMAN    16 T   LELETNGRD   0.432   CaM-II      .
# sp_P21980_TGM2_HUMAN    16 T   LELETNGRD   0.411   cdc2        .
# sp_P21980_TGM2_HUMAN    16 T   LELETNGRD   0.367   CKI         .
# sp_P21980_TGM2_HUMAN    16 T   LELETNGRD   0.342   DNAPK       .
# sp_P21980_TGM2_HUMAN    16 T   LELETNGRD   0.302   p38MAPK     .
# sp_P21980_TGM2_HUMAN    16 T   LELETNGRD   0.267   PKC         .
# sp_P21980_TGM2_HUMAN    16 T   LELETNGRD   0.240   ATM         .
# sp_P21980_TGM2_HUMAN    16 T   LELETNGRD   0.237   PKG         .
# sp_P21980_TGM2_HUMAN    16 T   LELETNGRD   0.192   RSK         .
# sp_P21980_TGM2_HUMAN    16 T   LELETNGRD   0.171   cdk5        .
# sp_P21980_TGM2_HUMAN    16 T   LELETNGRD   0.088   PKA         .
# sp_P21980_TGM2_HUMAN    16 T   LELETNGRD   0.085   PKB         .
# sp_P21980_TGM2_HUMAN    16 T   LELETNGRD   0.029   unsp        .
#
# sp_P21980_TGM2_HUMAN    23 T   RDHHTADLC   0.568   CKII       YES
# sp_P21980_TGM2_HUMAN    23 T   RDHHTADLC   0.452   CaM-II      .
# sp_P21980_TGM2_HUMAN    23 T   RDHHTADLC   0.448   GSK3        .
# sp_P21980_TGM2_HUMAN    23 T   RDHHTADLC   0.388   cdc2        .
# sp_P21980_TGM2_HUMAN    23 T   RDHHTADLC   0.361   CKI         .
# sp_P21980_TGM2_HUMAN    23 T   RDHHTADLC   0.343   PKG         .
# sp_P21980_TGM2_HUMAN    23 T   RDHHTADLC   0.342   DNAPK       .
# sp_P21980_TGM2_HUMAN    23 T   RDHHTADLC   0.325   p38MAPK     .
# sp_P21980_TGM2_HUMAN    23 T   RDHHTADLC   0.260   ATM         .
# sp_P21980_TGM2_HUMAN    23 T   RDHHTADLC   0.224   PKA         .
# sp_P21980_TGM2_HUMAN    23 T   RDHHTADLC   0.201   RSK         .
# sp_P21980_TGM2_HUMAN    23 T   RDHHTADLC   0.153   cdk5        .
# sp_P21980_TGM2_HUMAN    23 T   RDHHTADLC   0.090   PKC         .
```

**Fig4. Server output for the FASTA sequence showing amino acid prediction results**

```
%1  .....S.T....T....S....T.................T.........   #    100
%1  S....T...........S...ST.....S...................Y.   #    150
%1  .S......Y..T........S.............................   #    200
%1  ...........S..SS..Y.....S...................Y....S   #    250
%1  ..S..............................T........T...T.   #    300
%1  ..S.........................S.....................   #    350
%1  .........T....S...Y................ST..............   #    400
%1  ..........S...S...S.......S..S...........T......SS   #    450
%1  ..................................................T....   #    500
%1  ..Y...........S.............Y.........S..S.........   #    550
%1  .......S...............S........Y..................   #    600
%1  ......................T......T.....................   #    650
%1  ....................................
```

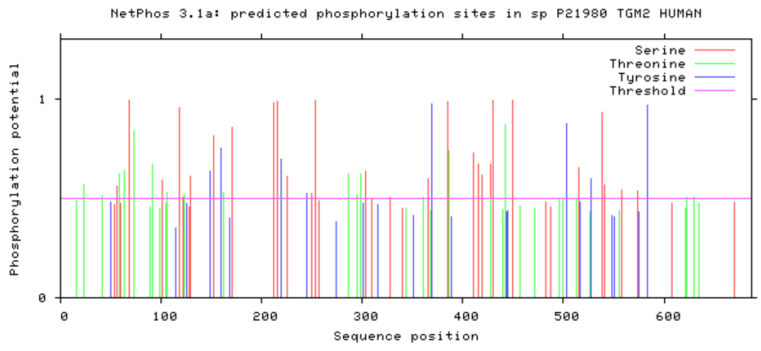



**Fig5. Predicted phosphorylation sites in the sequence according to NetPhos 3.1**

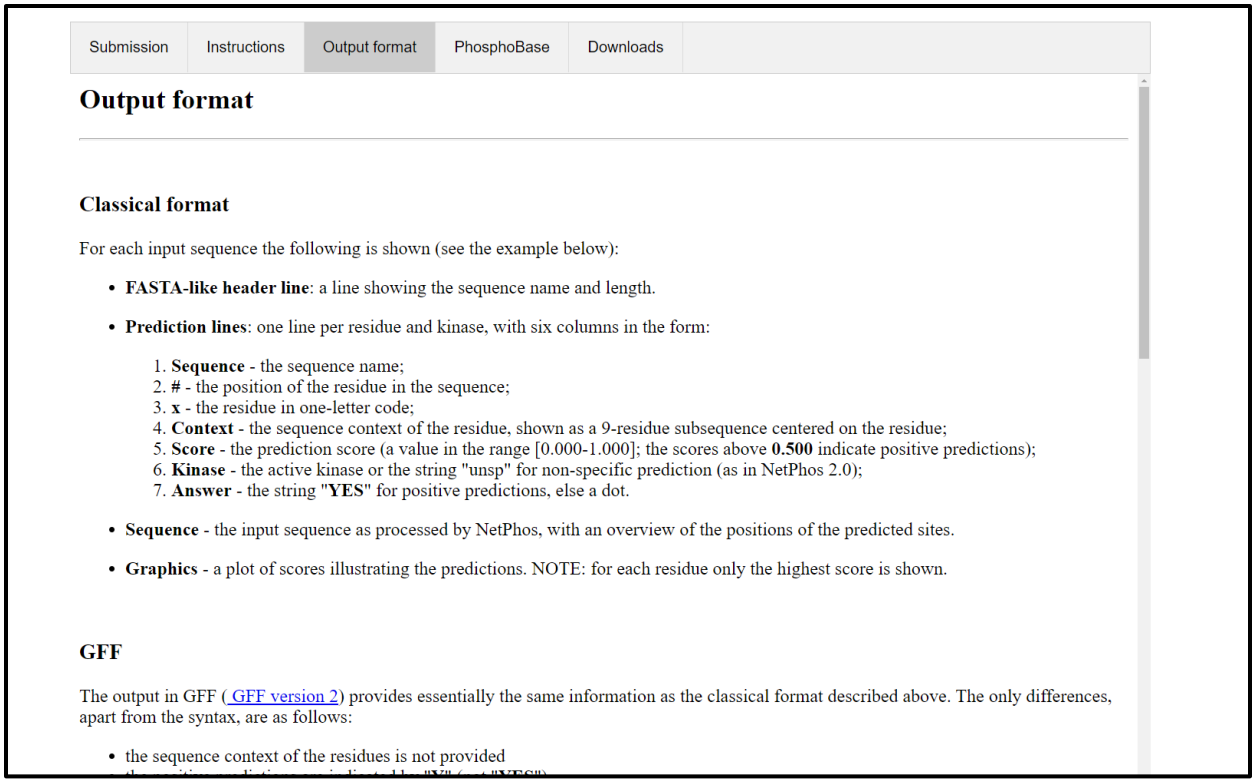



**Fig6. Output format of NetPhos 3.1 used to interpret results for query accordingly**

## Result:

- After submitting the result we can determine that there are 56 potential phosphorylation sites in the given protein sequence
- This interpretation was made on the basis that 56 results have a score about 0.5 and answer is YES
- The active kinase is also given and for non-specific prediction it is "**unsp**"

## Observation:

- NetPhos is a great tool to find the phosphorylation sites of a given protein sequence.
- Multiple sequences can be submitted at once and the results can be interpreted in bulk making it easier to interpret.

## References:


- Uniprot. (n.d.). Retrieved March 3, 2022, from https://www.uniprot.org/uniprot/P21980.fasta
- Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. Blom, N., Gammeltoft, S., and Brunak, S. Journal of Molecular Biology: 294(5): 1351-1362, 1999.
- Cruzat, V., Macedo Rogero, M., Noel Keane, K., Curi, R., & Newsholme, P. (2018). Glutamine: Metabolism and immune function, supplementation and clinical translation. *Nutrients*, *10*(11), 1564. https://doi.org/10.3390/nu10111564