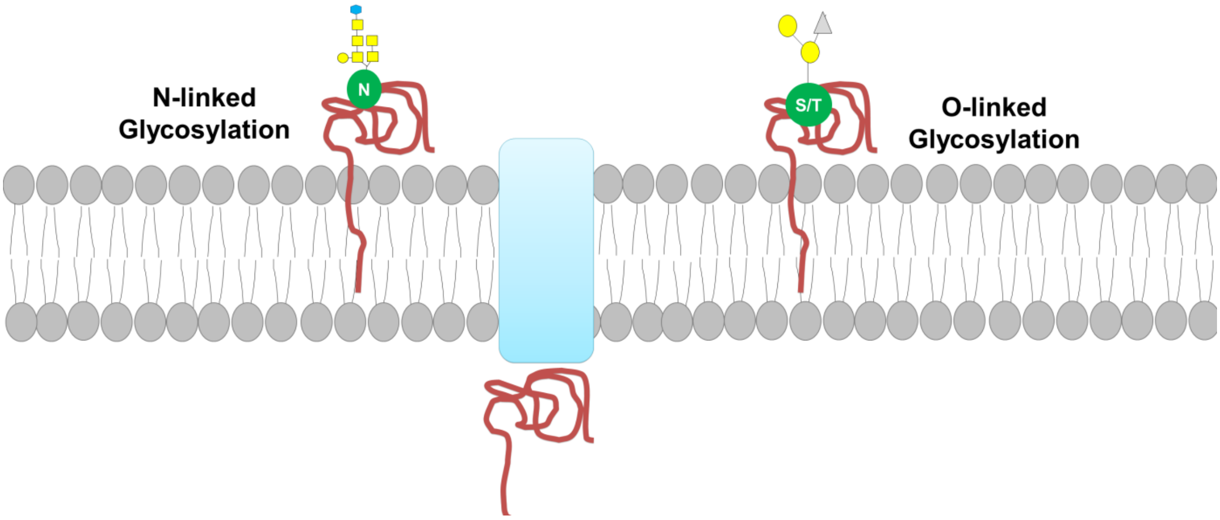
GlycoPP V2.0 FRAMEWORK



Bioinformatics Centre, CSIR-Institute of Microbial Technology,

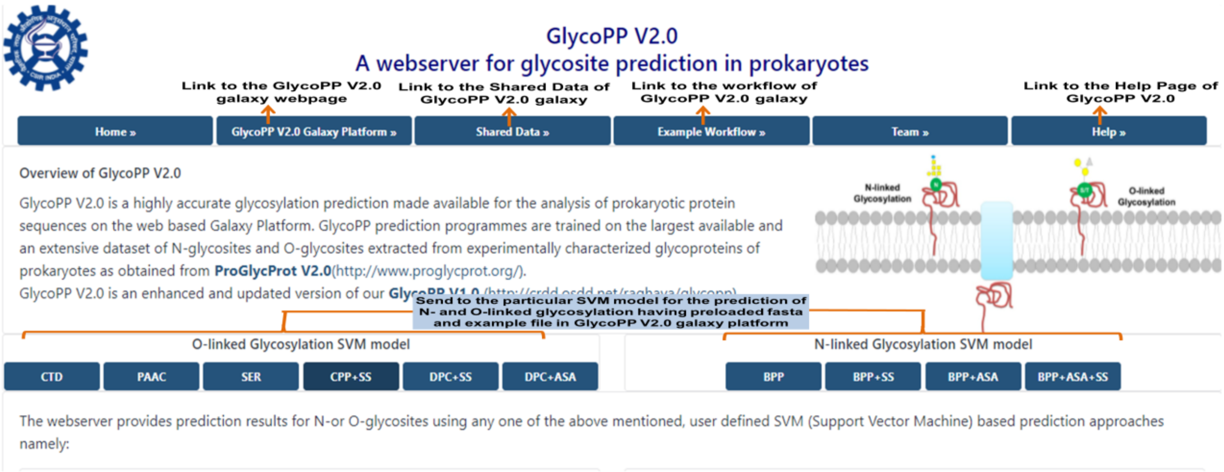
Chandigarh – 110036, India

Introduction

GlycoPP V2.0 is a highly accurate glycosylation prediction made available for the analysis of prokaryotic protein sequences on the web based Galaxy Platform. GlycoPP prediction programmes are trained on the largest available and an extensive dataset of N- and O-linked glycosites extracted from experimentally characterized glycoproteins of prokaryotes as obtained from ProGlycProt V2.0(http://www.proglycprot.org/).

GlycoPP V2.0 is an enhanced and updated version of our GlycoPP V1.0 (http://crdd.osdd.net/raghava/glycopp). The workflow system is implemented using the open source workflow architecture, Galaxy. GlycoPP2 is freely available and can be accessed at https://ab-openlab.csir.res.in/alkarao/glycopp2/#

There are three modules in GlycoPP2, described in this manual. All the modules are accessible without registering to the system. However, for maintaining user-sessions it is recommended that anyone who is interested in creating data or task intensive workflow should register (**Figure**). The benefit of registration includes user sessions, saved histories, visualization, generation and execution of workflow and many others.


GlycoPP V2.0
A webserver for glycosite prediction in prokaryotes
Link to the GlycoPP V2.0 galaxy webpage
Link to the Shared Data of GlycoPP V2.0 galaxy
Link to the workflow of GlycoPP V2.0 galaxy
Link to the Help Page of GlycoPP V2.0
Home »
GlycoPP V2.0 Galaxy Platform »
Shared Data »
Example Workflow »
Team »
Help »
N-linked Glycosylation
O-linked Glycosylation
Send to the particular SVM model for the prediction of N- and O-linked glycosylation having preloaded fasta and example file in GlycoPP V2.0 galaxy platform
O-linked Glycosylation SVM model
CTD
PAAC
SER
CPP+SS
DPC+SS
DPC+ASA
N-linked Glycosylation SVM model
BPP
BPP+SS
BPP+ASA
BPP+ASA+SS

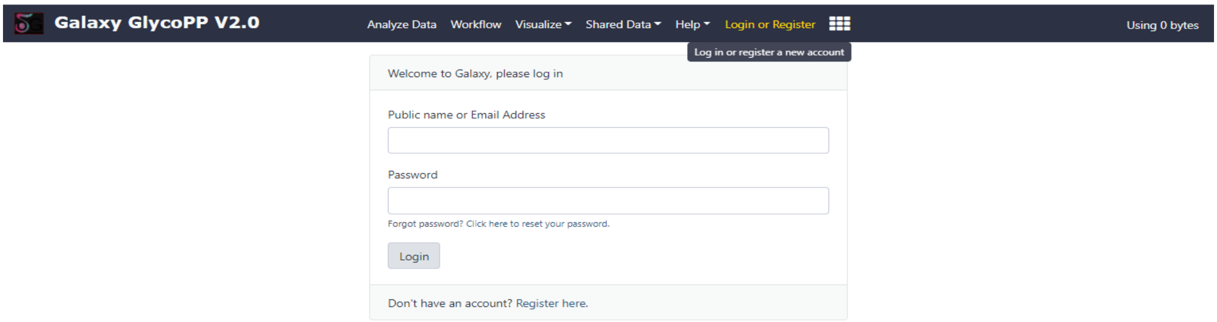
Overview of GlycoPP V2.0

GlycoPP V2.0 is a highly accurate glycosylation prediction made available for the analysis of prokaryotic protein sequences on the web based Galaxy Platform. GlycoPP prediction programmes are trained on the largest available and an extensive dataset of N-glycosites and O-glycosites extracted from experimentally characterized glycoproteins of prokaryotes as obtained from **ProGlycProt V2.0**(http://www.proglycprot.org/).

GlycoPP V2.0 is an enhanced and updated version of our **GlycoPP V1.0** (http://crdd.osdd.net/raghava/glycopp)

The webserver provides prediction results for N-or O-glycosites using any one of the above mentioned, user defined SVM (Support Vector Machine) based prediction approaches namely:

Figure 1: GlycoPP V2.0 Main page



Welcome to Galaxy, please log in

Public name or Email Address

Password

Forgot password? Click here to reset your password.

Don't have an account? Register here.

Figure 2: *GlycoPP2 Login Page*

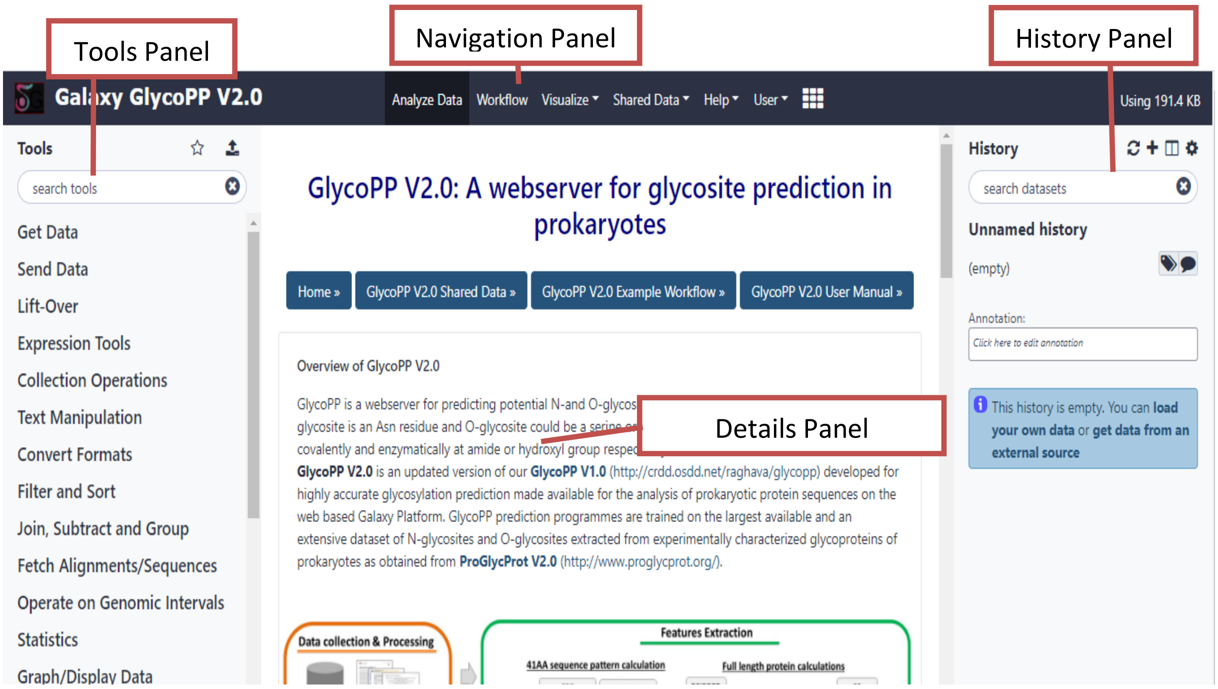
All the modules of GlycoPP2 and some default modules by Galaxy are accessible through a web-based interface which has following components (*Figure*).

Navigation Panel: It provides the links to major components of the server like Tools Page (Analyze Data), Workflow System, Shared Libraries, Visualization, Help Section and User Login/Registration.

Tool Panel: This panel lists all the tools available in GlycoPP2 along with default utilities in Galaxy.

Detail Panel (Canvas): This panel displays the interface of all the tools along with Input Parameters required to run a tool. It also provides help and examples to run a tool. This panel also displays the Output of a tool after its execution when user clicks on the eye 👁 icon show in **History Panel**.

History Panel: This panel shows the information about the tools which are executed by a user. The information can include result after completion of a tool execution or error generated while running the tool. The workflow(s) are generated by extracting tasks from history panel.



GlycoPP V2.0: A webserver for glycosite prediction in prokaryotes

Overview of GlycoPP V2.0

GlycoPP is a webserver for predicting potential N-and O-glycos glycosite is an Asn residue and O-glycosite could be a serine or covalently and enzymatically at amide or hydroxyl group respec **GlycoPP V2.0** is an updated version of our **GlycoPP V1.0** (http://crdd.osdd.net/raghava/glycopp) developed for highly accurate glycosylation prediction made available for the analysis of prokaryotic protein sequences on the web based Galaxy Platform. GlycoPP prediction programmes are trained on the largest available and an extensive dataset of N-glycosites and O-glycosites extracted from experimentally characterized glycoproteins of prokaryotes as obtained from **ProGlycProt V2.0** (http://www.proglycprot.org/).

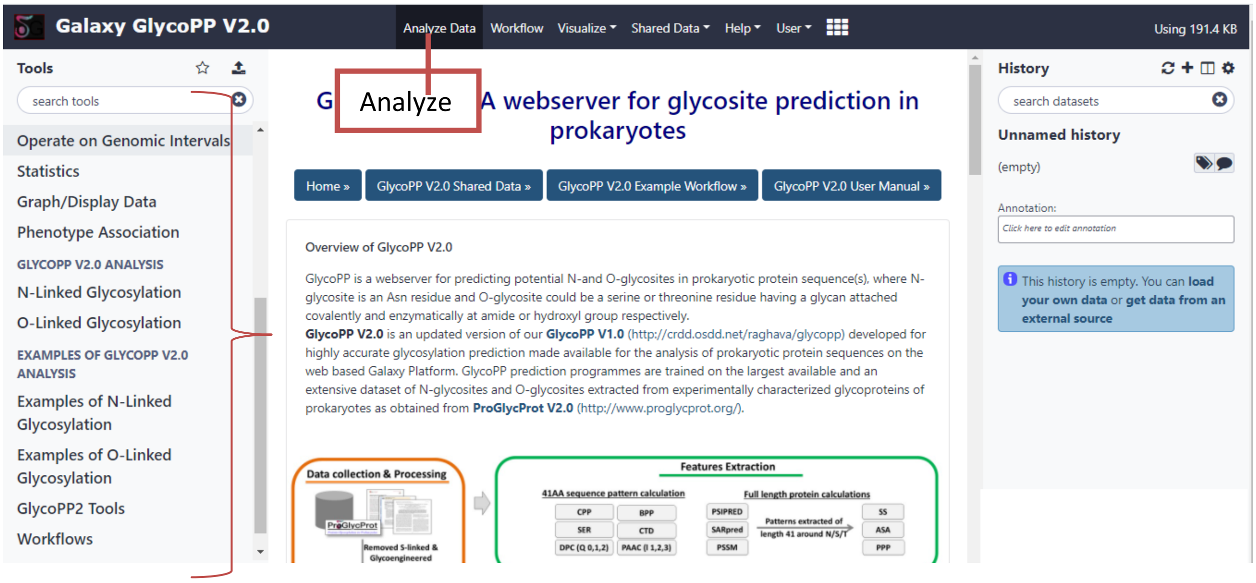
This history is empty. You can **load your own data** or **get data from an external source**

Figure 3: *GlycoPP2 Homepage*

I. Navigation Panel

1. Analyze Data

The data analysis page is where everything happens. There, you can run any available tools on the data, run complete workflows, browse or download a result, and share files with other users. It is the default page when you open Galaxy in your browser, but you can also access it any time by clicking on "Analyze Data" in the Navigation Panel.


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
User
Using 191.4 KB
Tools
search tools
Operate on Genomic Intervals
Statistics
Graph/Display Data
Phenotype Association
GLYCOPP V2.0 ANALYSIS
N-Linked Glycosylation
O-Linked Glycosylation
EXAMPLES OF GLYCOPP V2.0 ANALYSIS
Examples of N-Linked Glycosylation
Examples of O-Linked Glycosylation
GlycoPP2 Tools
Workflows
Analyze
G A webserver for glycosite prediction in prokaryotes
Home »
GlycoPP V2.0 Shared Data »
GlycoPP V2.0 Example Workflow »
GlycoPP V2.0 User Manual »
History
search datasets
Unnamed history
(empty)
Annotation:
Click here to edit annotation
This history is empty. You can load your own data or get data from an external source
Data collection & Processing
ProGlycProt
Removed S-linked & Glycoengineered
Features Extraction
41AA sequence pattern calculation
CPP
BPP
SER
CTD
DPC (Q 0,1,2)
PAAC (I 1,2,3)
Full length protein calculations
PSIPRED
SARpred
PSSM
Patterns extracted of length 41 around N/S/T
SS
ASA
PPP

Overview of GlycoPP V2.0

GlycoPP is a webserver for predicting potential N-and O-glycosites in prokaryotic protein sequence(s), where N-glycosite is an Asn residue and O-glycosite could be a serine or threonine residue having a glycan attached covalently and enzymatically at amide or hydroxyl group respectively.
GlycoPP V2.0 is an updated version of our **GlycoPP V1.0** (http://crdd.osdd.net/raghava/glycopp) developed for highly accurate glycosylation prediction made available for the analysis of prokaryotic protein sequences on the web based Galaxy Platform. GlycoPP prediction programmes are trained on the largest available and an extensive dataset of N-glycosites and O-glycosites extracted from experimentally characterized glycoproteins of prokaryotes as obtained from **ProGlycProt V2.0** (http://www.proglycprot.org/).

2. Workflow

Workflows are analyses that are intended to be executed (one or more times) with different user-provided input Datasets. Workflow can be reused over and over, not only reducing tedious work, but enhancing reproducibility by applying the same exact methods to all of your data. Workflow is nothing but creating pipeline, user can use it again and again or user can published it.

Workflow can be created through navigation panel or from tool panel. In workflow section user can create workflow or can upload or import the workflow. The canvas is where inputs, tools, and noodles are added and connected as you build and modify your workflow (Figure 4). Selecting Edit opens the workflow editor view (Figure 3). The navigator provides a full view of your workflow in a condensed format (Figure 4). Accessed by clicking on the gear icon on the right side of the center Workflow Canvas upper bar, the workflow editor menu (Figure 3) is for global editor actions. It consists of Save, Run, Edit Attributes, Auto re-layout, Close

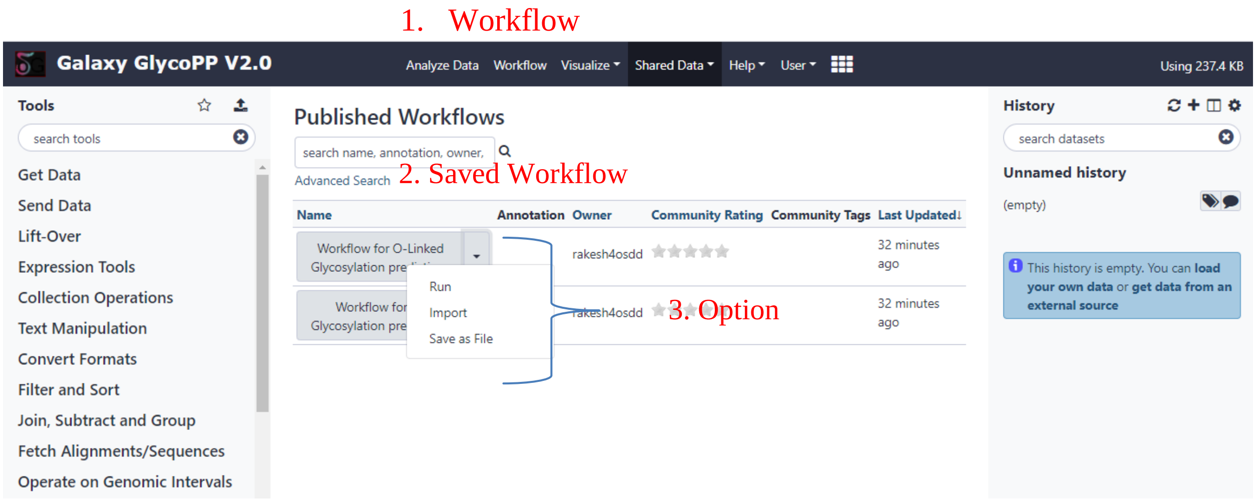


Figure 5: *Options for workflow*

The following example of workflow shows the “Prediction of N- and O-linked glycosylation prediction”. The prediction workflow can use the four implemented svm model for N-linked glycosyation and six svm model for O-linked glycosylation

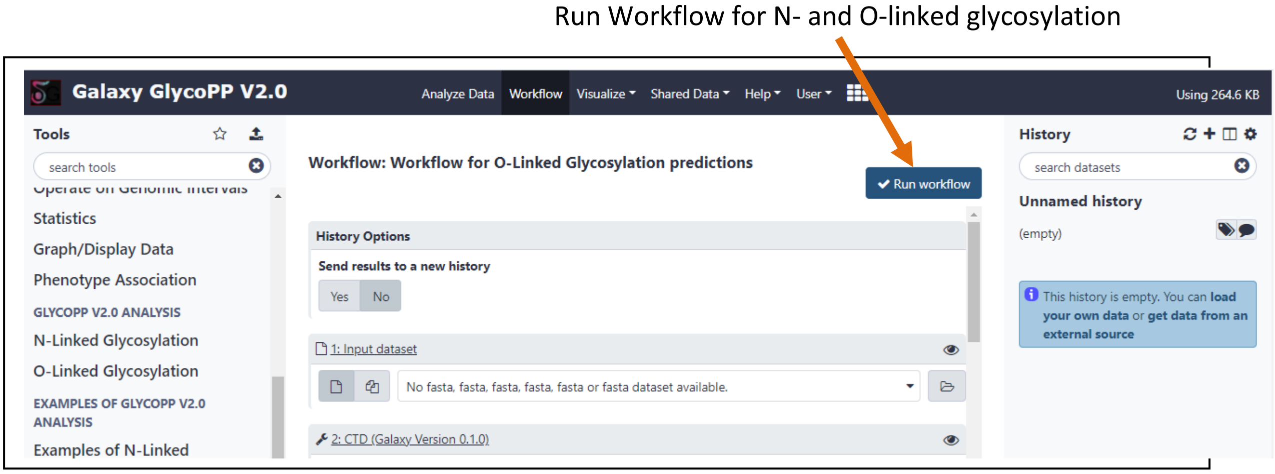

Run Workflow for N- and O-linked glycosylation
Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
User
Using 264.6 KB
Tools
search tools
Operate on Genomic Intervals
Statistics
Graph/Display Data
Phenotype Association
GLYCOPP V2.0 ANALYSIS
N-Linked Glycosylation
O-Linked Glycosylation
EXAMPLES OF GLYCOPP V2.0 ANALYSIS
Examples of N-Linked
Workflow: Workflow for O-Linked Glycosylation predictions
✓ Run workflow
History Options
Send results to a new history
Yes
No
1: Input dataset
No fasta, fasta, fasta, fasta, fasta or fasta dataset available.
2: CTD (Galaxy Version 0.1.0)
History
search datasets
Unnamed history
(empty)
This history is empty. You can load your own data or get data from an external source

Figure 6: *Workflow overview*

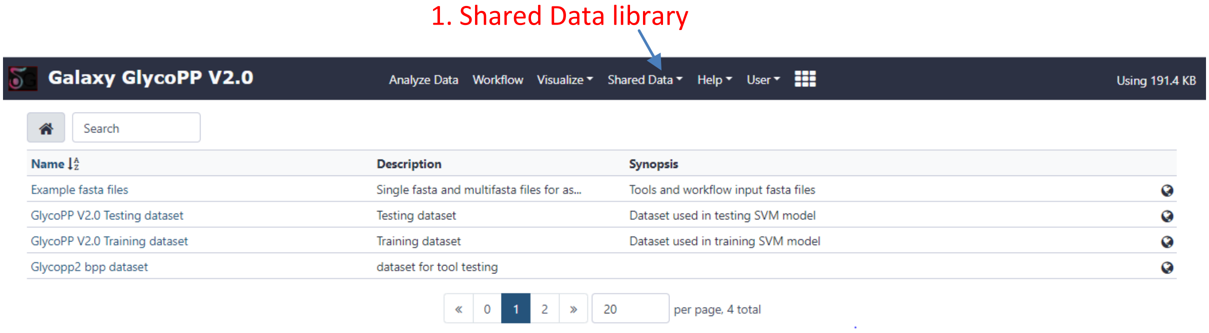
3. Shared data library

Data libraries are collections of Datasets that are accessible from within a Galaxy instance. Libraries are designed for sharing datasets in between users or groups. The data library of GlycoPP2 consists of prokaryotic glycoproteins list protein list used in SVM model generation. Some of the actions that can be performed on data libraries are accessed by clicking the pop-up menu icon just right of the data library name.

- View Information –Shows the information about dataset.
- Import this dataset into your current history - this creates an item in your current

history on which you can perform analysis. The item is a pointer to the library dataset disk file, so the file is not copied on disk.

- Download this dataset - this allows you to download a local copy of the dataset.



1. Shared Data library

Name	Description	Synopsis
Example fasta files	Single fasta and multifasta files for as...	Tools and workflow input fasta files
GlycoPP V2.0 Testing dataset	Testing dataset	Dataset used in testing SVM model
GlycoPP V2.0 Training dataset	Training dataset	Dataset used in training SVM model
Glycopp2 bpp dataset	dataset for tool testing	

Figure 6: Viewed for shared library

4 Help

The help section of Galaxy consists of Support, Search, Mailing List, Videos, Wiki and How to cite Galaxy. User can fine user manual in help section.

5. User

Login option and register option can get in user section. It is recommended that user register their account before using framework. Although unregistered users have access to tools available but their history is stored temporarily. On the other hand, registered users can save and retrieve their results in history panel later too.

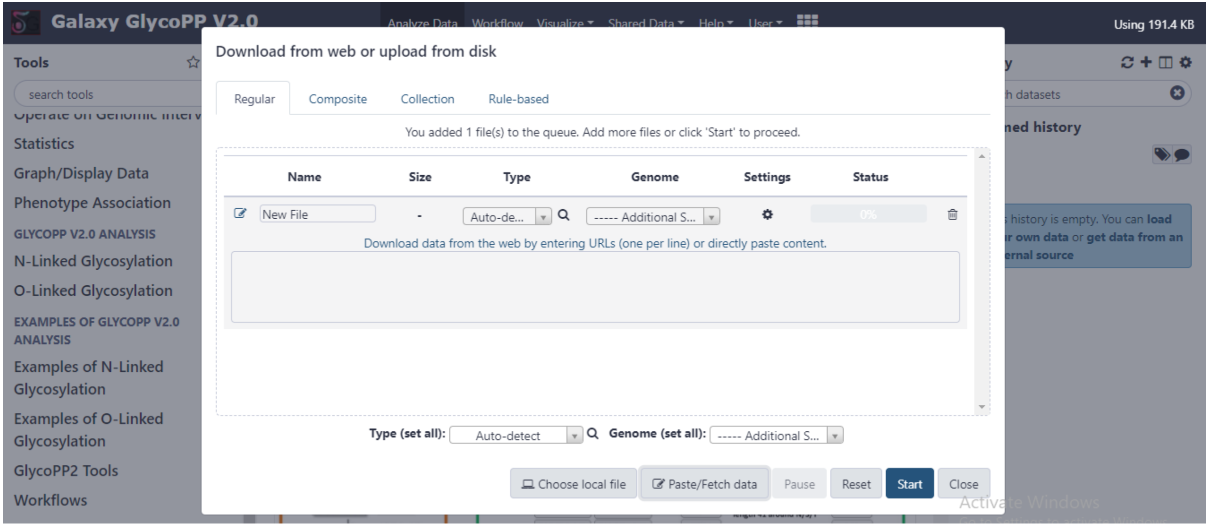
Tool Panel – GlycoPP2 tools

The user can get tools in tool panel. There are two categories of tools galaxy inbuilt tool and GlycoPP2 tools. Galaxy tools consists of Data importing, Manipulation, Filtering, Sorting, Format conversion etc. GlycoPP2 tool are specific for finding N- and O-linked glycosite in prokaryotic protein sequence

1. Importing data to the GlycoPP2

A user can upload the data using the Galaxy tool **Get Data.** The uploaded data can be used for GlycoPP V2.0 Analysis.The following figure shows the file upload method.

1. Get data / Upload file



You added 1 file(s) to the queue. Add more files or click 'Start' to proceed.

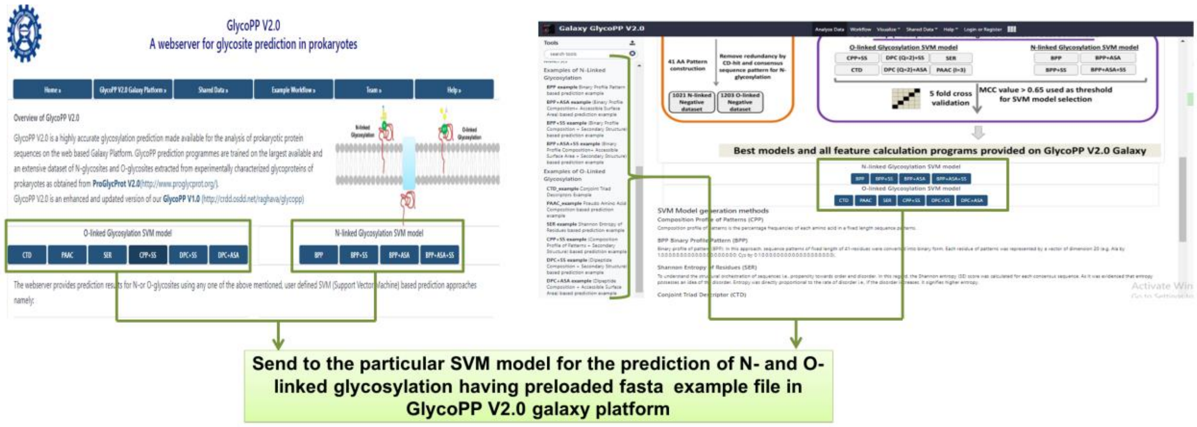
Download data from the web by entering URLs (one per line) or directly paste content.

history is empty. You can load r own data or get data from an ernal source

Figure 8: *Get Data screen*

2. Examples of Glycopp V2.0 Analysis

These have example fasta and multifasta file reloded for the prediction analysis of N- and O-linked glycosylation SVM model. BPP, BPP+ASA, BPP+SS and BPP+ASA+SS for N-linked Glycosylation and CTD, PAAC, SER, CPP+SS, DPC+SS and DPC+ASA for O-linked Glycosylation


GlycoPP V2.0
A webserver for glycosite prediction in prokaryotes
Home
GlycoPP V2.0 Galaxy Platform
Shared Data
Example Workflow
Team
Help
Overview of GlycoPP V2.0
N-linked Glycosylation
O-linked Glycosylation
O-linked Glycosylation SVM model
CTD
PAAC
SER
CPP+SS
DPC+SS
DPC+ASA
N-linked Glycosylation SVM model
BPP
BPP+SS
BPP+ASA
BPP+ASA+SS
Galaxy GlycoPP V2.0
Tools
Examples of N-Linked Glycosylation
Examples of O-Linked Glycosylation
41 AA Pattern construction
Remove redundancy by CD-hit and consensus sequence pattern for N-glycosylation
1021 N-linked Negative dataset
1203 O-linked Negative dataset
O-linked Glycosylation SVM model
CPP+SS
DPC (G=2)+SS
SER
CTD
DPC (G=2)+ASA
PAAC (λ=3)
N-linked Glycosylation SVM model
BPP
BPP+ASA
BPP+SS
BPP+ASA+SS
5 fold cross validation
MCC value > 0.65 used as threshold for SVM model selection
Best models and all feature calculation programs provided on GlycoPP V2.0 Galaxy
SVM Model generation methods
Composition Profile of Patterns (CPP)
BPP Binary Profile Pattern (BPP)
Shannon Entropy of Residues (SER)
Conjoint Triad Descriptor (CTD)
Activate Win

GlycoPP V2.0 is a highly accurate glycosylation prediction made available for the analysis of prokaryotic protein sequences on the web based Galaxy Platform. GlycoPP prediction programmes are trained on the largest available and an extensive dataset of N-glycosites and O-glycosites extracted from experimentally characterized glycoproteins of prokaryotes as obtained from ProGlycProt V2.0 (http://www.proglycprot.org/).

GlycoPP V2.0 is an enhanced and updated version of our GlycoPP V1.0 (http://crdd.osdd.net/raghava/glycopp)

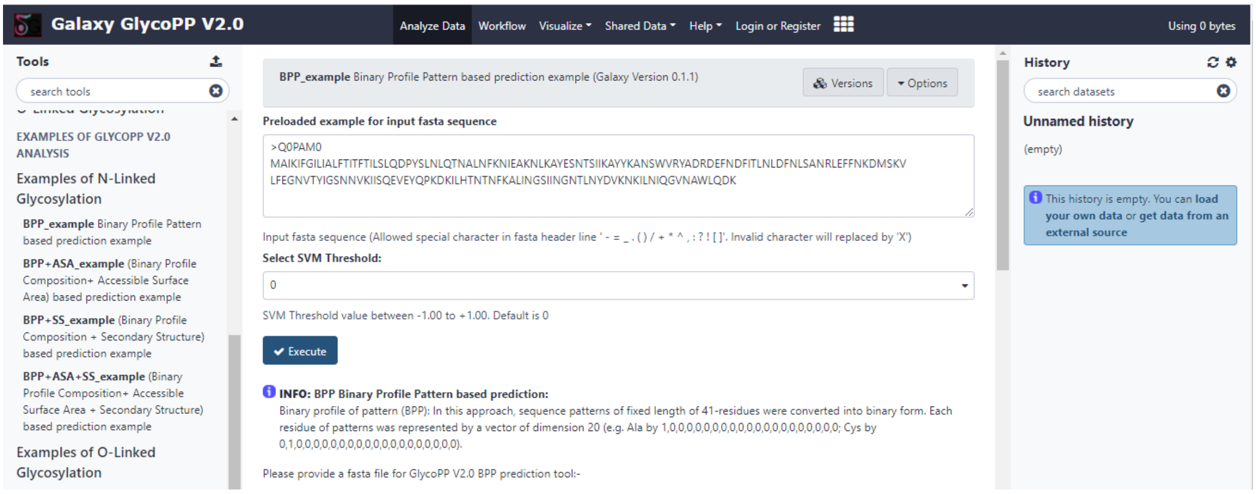
The webserver provides prediction results for N-or O-glycosites using any one of the above mentioned, user defined SVM (Support Vector Machine) based prediction approaches namely:

Send to the particular SVM model for the prediction of N- and O-linked glycosylation having preloaded fasta example file in GlycoPP V2.0 galaxy platform

Figure 9: *Example of GlycoPP V2.0 Analsis Screen*

a) Example of N-linked Glycosylation

BPP example Binary Profile Pattern based prediction example


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 0 bytes
Tools
search tools
EXAMPLES OF GLYCOPP V2.0 ANALYSIS
Examples of N-Linked Glycosylation
BPP_example Binary Profile Pattern based prediction example
BPP+ASA_example (Binary Profile Composition+ Accessible Surface Area) based prediction example
BPP+SS_example (Binary Profile Composition + Secondary Structure) based prediction example
BPP+ASA+SS_example (Binary Profile Composition+ Accessible Surface Area + Secondary Structure) based prediction example
Examples of O-Linked Glycosylation
Versions
Options
Execute
History
search datasets
Unnamed history
(empty)

BPP_example Binary Profile Pattern based prediction example (Galaxy Version 0.1.1)

Preloaded example for input fasta sequence

```
>Q0PAM0
MAIKIFGILIALFTITFTILSLQDPYSLNLQTNALNFKNIEAKNLKAYESNTSIIKAYYKANSWVRYADRDEFNDFITLNLDFNLSANRLEFFNKDMSKV
LFEGNVTYIGSNNVKIISQEVEYQPKDKILHTNTNFKALINGSIINGNTLNYDVKNKILNIQGVNAWLQDK
```

Input fasta sequence (Allowed special character in fasta header line ' - = _ . ( ) / + * ^ , : ? ! [ ]'. Invalid character will replaced by 'X')

Select SVM Threshold:

0

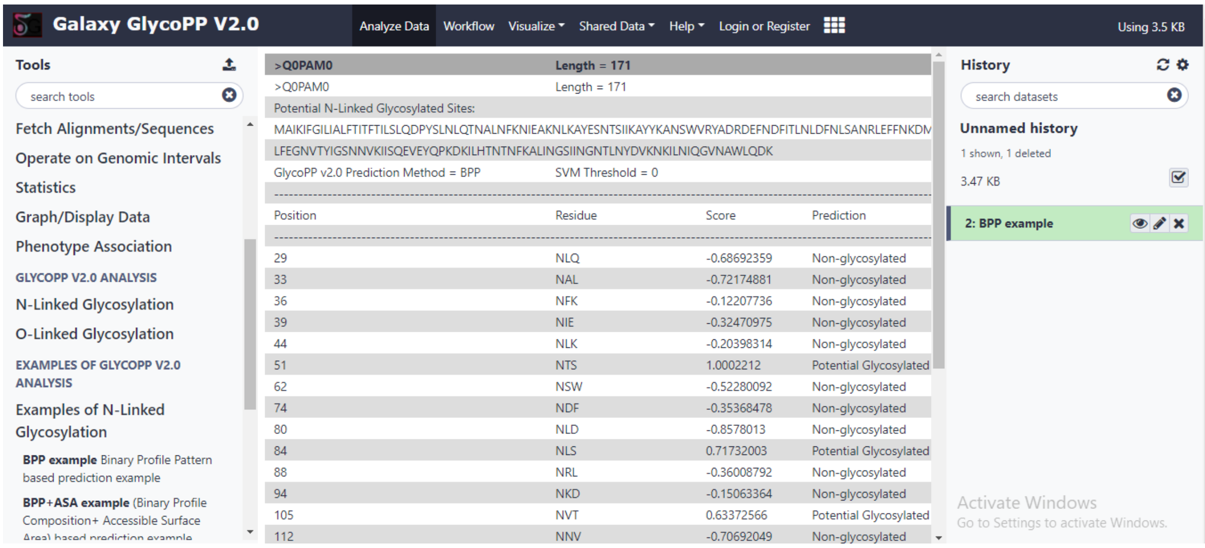
SVM Threshold value between -1.00 to +1.00. Default is 0

INFO: BPP Binary Profile Pattern based prediction:
Binary profile of pattern (BPP): In this approach, sequence patterns of fixed length of 41-residues were converted into binary form. Each residue of patterns was represented by a vector of dimension 20 (e.g. Ala by 1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0; Cys by 0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0).

Please provide a fasta file for GlycoPP V2.0 BPP prediction tool:-

This history is empty. You can **load your own data** or **get data from an external source**

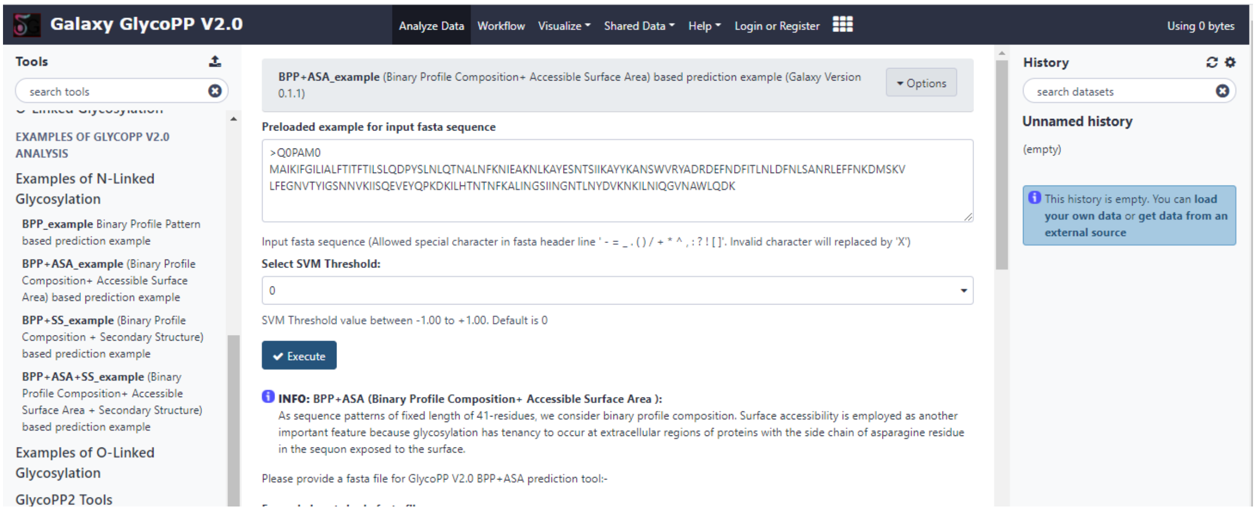
Figure-10 : *Example of GlycoPP V2.0 Analsis overview*



```
>Q0PAM0                                 Length = 171
>Q0PAM0                                 Length = 171
Potential N-Linked Glycosylated Sites:
MAIKIFGILIALFTITFTILSLQDPYSLNLQTNALNFKNIEAKNLKAYESNTSIIKAYYKANSWVRYADRDEFNDFITLNLDFNLSANRLEFFNKDM
LFEGNVTYIGSNNVKIISQEVEYQPKDKILHTNTNFKALINGSIINGNTLNYDVKNKILNIQGVNAWLQDK
GlycoPP v2.0 Prediction Method = BPP    SVM Threshold = 0
------------------------------------------------------------------------------------------
Position                                Residue         Score           Prediction
------------------------------------------------------------------------------------------
29                                      NLQ             -0.68692359     Non-glycosylated
33                                      NAL             -0.72174881     Non-glycosylated
36                                      NFK             -0.12207736     Non-glycosylated
39                                      NIE             -0.32470975     Non-glycosylated
44                                      NLK             -0.20398314     Non-glycosylated
51                                      NTS             1.0002212       Potential Glycosylated
62                                      NSW             -0.52280092     Non-glycosylated
74                                      NDF             -0.35368478     Non-glycosylated
80                                      NLD             -0.8578013      Non-glycosylated
84                                      NLS             0.71732003      Potential Glycosylated
88                                      NRL             -0.36008792     Non-glycosylated
94                                      NKD             -0.15063364     Non-glycosylated
105                                     NVT             0.63372566      Potential Glycosylated
112                                     NNV             -0.70692049     Non-glycosylated
```

Figure11: *Example of GlycoPP V2.0 Analsis output for BPP (Binary Profile Pattern) forN-linked glycosylation*

BPP+ASA example (Binary Profile Composition+ Accessible Surface Area) based prediction example


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 0 bytes
Tools
search tools
EXAMPLES OF GLYCOPP V2.0 ANALYSIS
Examples of N-Linked Glycosylation
BPP_example Binary Profile Pattern based prediction example
BPP+ASA_example (Binary Profile Composition+ Accessible Surface Area) based prediction example
BPP+SS_example (Binary Profile Composition + Secondary Structure) based prediction example
BPP+ASA+SS_example (Binary Profile Composition+ Accessible Surface Area + Secondary Structure) based prediction example
Examples of O-Linked Glycosylation
GlycoPP2 Tools
Options
Execute
History
search datasets
Unnamed history
(empty)
This history is empty. You can load your own data or get data from an external source

BPP+ASA_example (Binary Profile Composition+ Accessible Surface Area) based prediction example (Galaxy Version 0.1.1)

Preloaded example for input fasta sequence

```
>Q0PAM0
MAIKIFGILIALFTITFTILSLQDPYSLNLQTNALNFKNIEAKNLKAYESNTSIIKAYYKANSWVRYADRDEFNDFITLNLDFNLSANRLEFFNKDMSKV
LFEGNVTYIGSNNVKIISQEVEYQPKDKILHTNTNFKALINGSIINGNTLNYDVKNKILNIQGVNAWLQDK
```

Input fasta sequence (Allowed special character in fasta header line ' - = _ . ( ) / + * ^ , : ? ! [ ]'. Invalid character will replaced by 'X')

Select SVM Threshold:

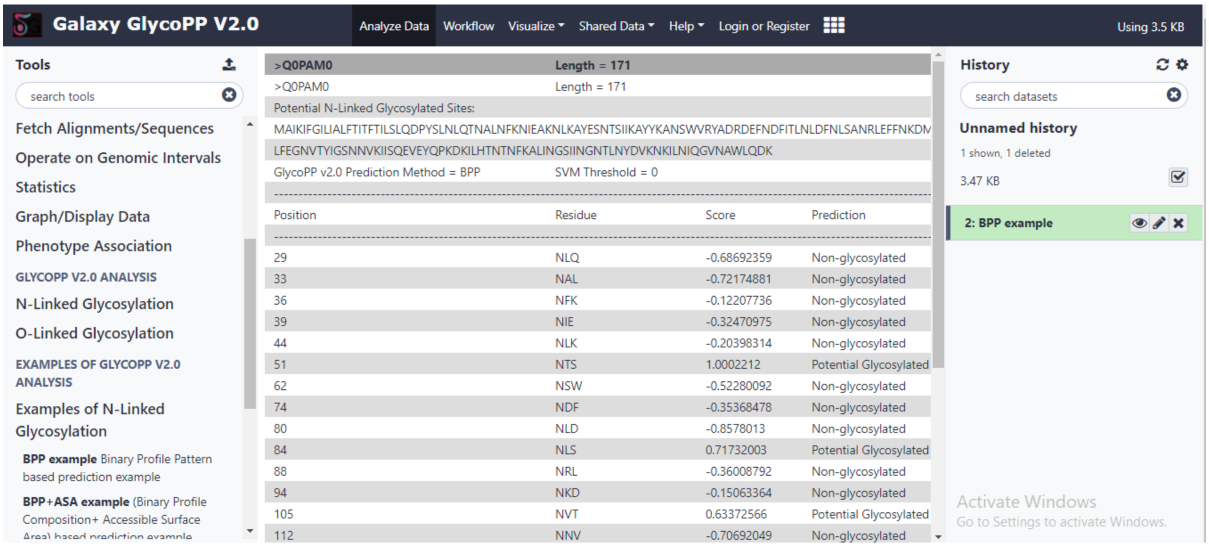
0

SVM Threshold value between -1.00 to +1.00. Default is 0

INFO: **BPP+ASA (Binary Profile Composition+ Accessible Surface Area):**
As sequence patterns of fixed length of 41-residues, we consider binary profile composition. Surface accessibility is employed as another important feature because glycosylation has tenancy to occur at extracellular regions of proteins with the side chain of asparagine residue in the sequon exposed to the surface.

Please provide a fasta file for GlycoPP V2.0 BPP+ASA prediction tool:-

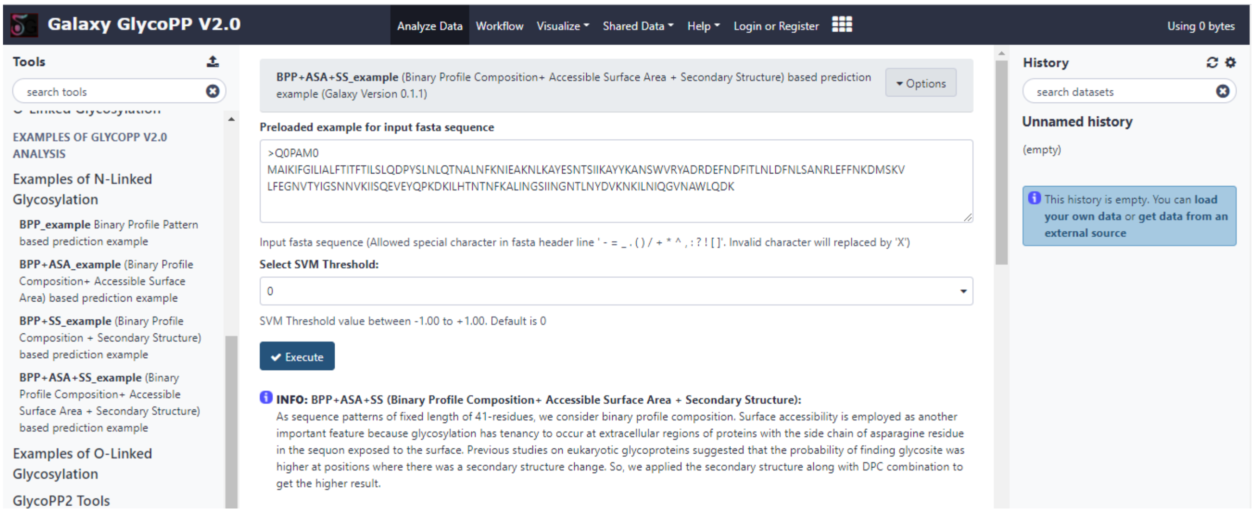
Figure 12: *Example of GlycoPP V2.0 Analsis overview*



>Q0PAM0	**Length = 171**		
>Q0PAM0	Length = 171		
Potential N-Linked Glycosylated Sites:			
MAIKIFGILIALFTITFTILSLQDPYSLNLQTNALNFKNIEAKNLKAYESNTSIIKAYYKANSWVRYADRDEFNDFITLNLDFNLSANRLEFFNKDM			
LFEGNVTYIGSNNVKIISQEVEYQPKDKILHTNTNFKALINGSIINGNTLNYDVKNKILNIQGVNAWLQDK			
GlycoPP v2.0 Prediction Method = BPP	SVM Threshold = 0		
--			
Position	Residue	Score	Prediction
--			
29	NLQ	-0.68692359	Non-glycosylated
33	NAL	-0.72174881	Non-glycosylated
36	NFK	-0.12207736	Non-glycosylated
39	NIE	-0.32470975	Non-glycosylated
44	NLK	-0.20398314	Non-glycosylated
51	NTS	1.0002212	Potential Glycosylated
62	NSW	-0.52280092	Non-glycosylated
74	NDF	-0.35368478	Non-glycosylated
80	NLD	-0.8578013	Non-glycosylated
84	NLS	0.71732003	Potential Glycosylated
88	NRL	-0.36008792	Non-glycosylated
94	NKD	-0.15063364	Non-glycosylated
105	NVT	0.63372566	Potential Glycosylated
112	NNV	-0.70692049	Non-glycosylated

Figure 13: Example of GlycoPP V2.0 Analsis output for BPP+ASA (Binary Profile Pattern+ Accessible Surface Are) forN-linked glycosylation

BPP+ASA+SS example (Binary Profile Composition+ Accessible Surface Area + Secondary Structure) based prediction example


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 0 bytes
Tools
search tools
EXAMPLES OF GLYCOPP V2.0 ANALYSIS
Examples of N-Linked Glycosylation
BPP_example Binary Profile Pattern based prediction example
BPP+ASA_example (Binary Profile Composition+ Accessible Surface Area) based prediction example
BPP+SS_example (Binary Profile Composition + Secondary Structure) based prediction example
BPP+ASA+SS_example (Binary Profile Composition+ Accessible Surface Area + Secondary Structure) based prediction example
Examples of O-Linked Glycosylation
GlycoPP2 Tools
Options
Execute
History
search datasets
Unnamed history
(empty)

BPP+ASA+SS_example (Binary Profile Composition+ Accessible Surface Area + Secondary Structure) based prediction example (Galaxy Version 0.1.1)

Preloaded example for input fasta sequence

```
>Q0PAM0
MAIKIFGILIALFTITFTILSLQDPYSLNLQTNALNFKNIEAKNLKAYESNTSIIKAYYKANSWVRYADRDEFNDFITLNLDFNLSANRLEFFNKDMSKV
LFEGNVTYIGSNNVKIISQEVEYQPKDKILHTNTNFKALINGSIINGNTLNYDVKNKILNIQGVNAWLQDK
```

Input fasta sequence (Allowed special character in fasta header line ' - = _ . ( ) / + * ^ , : ? ! [ ]'. Invalid character will replaced by 'X')

Select SVM Threshold:

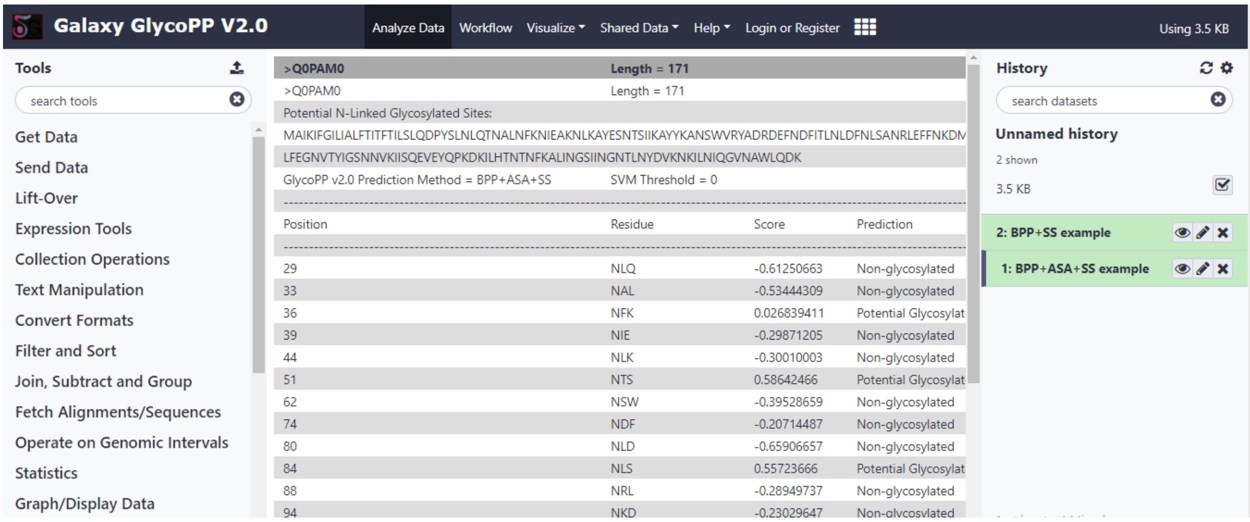
0

SVM Threshold value between -1.00 to +1.00. Default is 0

INFO: BPP+ASA+SS (Binary Profile Composition+ Accessible Surface Area + Secondary Structure):
As sequence patterns of fixed length of 41-residues, we consider binary profile composition. Surface accessibility is employed as another important feature because glycosylation has tenancy to occur at extracellular regions of proteins with the side chain of asparagine residue in the sequon exposed to the surface. Previous studies on eukaryotic glycoproteins suggested that the probability of finding glycosite was higher at positions where there was a secondary structure change. So, we applied the secondary structure along with DPC combination to get the higher result.

This history is empty. You can **load your own data** or **get data from an external source**

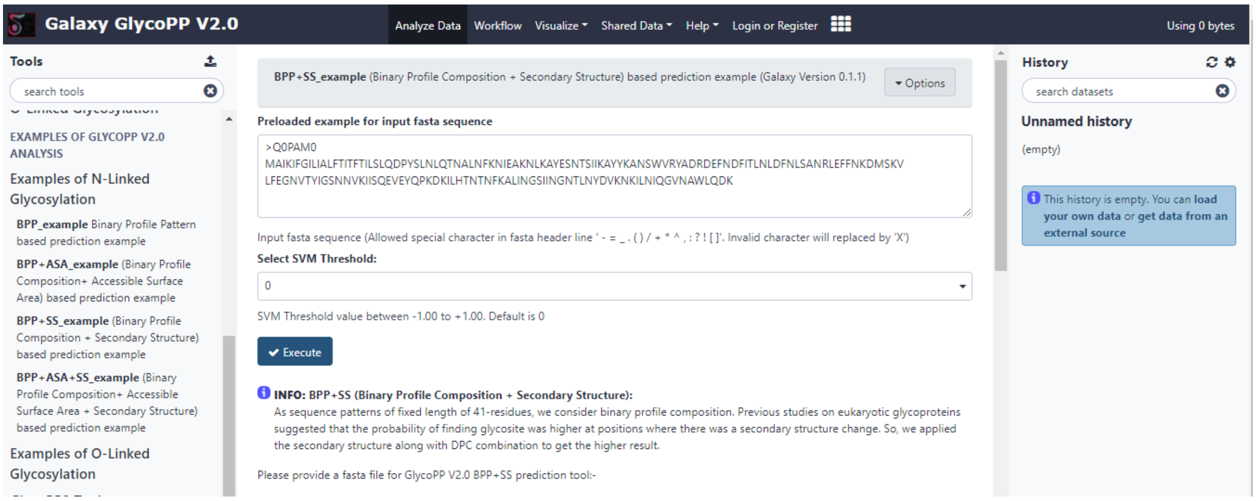
Figure 14: *Example of GlycoPP V2.0 Analsis overview*



```
>Q0PAM0                                          Length = 171
>Q0PAM0                                          Length = 171
Potential N-Linked Glycosylated Sites:
MAIKIFGILIALFTITFTILSLQDPYSLNLQTNALNFKNIEAKNLKAYESNTSIIKAYYKANSWVRYADRDEFNDFITLNLDFNLSANRLEFFNKDM
LFEGNVTYIGSNNVKIISQEVEYQPKDKILHTNTNFKALINGSIINGNTLNYDVKNKILNIQGVNAWLQDK
GlycoPP v2.0 Prediction Method = BPP+ASA+SS      SVM Threshold = 0
-------------------------------------------------------------------------------------------------
Position                                         Residue     Score          Prediction
-------------------------------------------------------------------------------------------------
29                                               NLQ         -0.61250663    Non-glycosylated
33                                               NAL         -0.53444309    Non-glycosylated
36                                               NFK         0.026839411    Potential Glycosylat
39                                               NIE         -0.29871205    Non-glycosylated
44                                               NLK         -0.30010003    Non-glycosylated
51                                               NTS         0.58642466     Potential Glycosylat
62                                               NSW         -0.39528659    Non-glycosylated
74                                               NDF         -0.20714487    Non-glycosylated
80                                               NLD         -0.65906657    Non-glycosylated
84                                               NLS         0.55723666     Potential Glycosylat
88                                               NRL         -0.28949737    Non-glycosylated
94                                               NKD         -0.23029647    Non-glycosylated
```

Figure 15: Example of GlycoPP V2.0 Analsis output for BPP+ASA+SS example (Binary Profile Composition+ Accessible Surface Area + Secondary Structure) forN-linked glycosylation

BPP+SS example (Binary Profile Composition+ Secondary Structure) based prediction example


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 0 bytes
Tools
search tools
O-Linked Glycosylation
EXAMPLES OF GLYCOPP V2.0 ANALYSIS
Examples of N-Linked Glycosylation
BPP_example Binary Profile Pattern based prediction example
BPP+ASA_example (Binary Profile Composition+ Accessible Surface Area) based prediction example
BPP+SS_example (Binary Profile Composition + Secondary Structure) based prediction example
BPP+ASA+SS_example (Binary Profile Composition+ Accessible Surface Area + Secondary Structure) based prediction example
Examples of O-Linked Glycosylation
Options
Execute
History
search datasets
Unnamed history
(empty)

BPP+SS_example (Binary Profile Composition + Secondary Structure) based prediction example (Galaxy Version 0.1.1)

Preloaded example for input fasta sequence

```
>Q0PAM0
MAIKIFGILIALFTITFTILSLQDPYSLNLQTNALNFKNIEAKNLKAYESNTSIIKAYYKANSWVRYADRDEFNDFITLNLDFNLSANRLEFFNKDMSKV
LFEGNVTYIGSNNVKIISQEVEYQPKDKILHTNTNFKALINGSIINGNTLNYDVKNKILNIQGVNAWLQDK
```

Input fasta sequence (Allowed special character in fasta header line ' - = _ . ( ) / + * ^ , : ? ! [ ]'. Invalid character will replaced by 'X')

Select SVM Threshold:

0

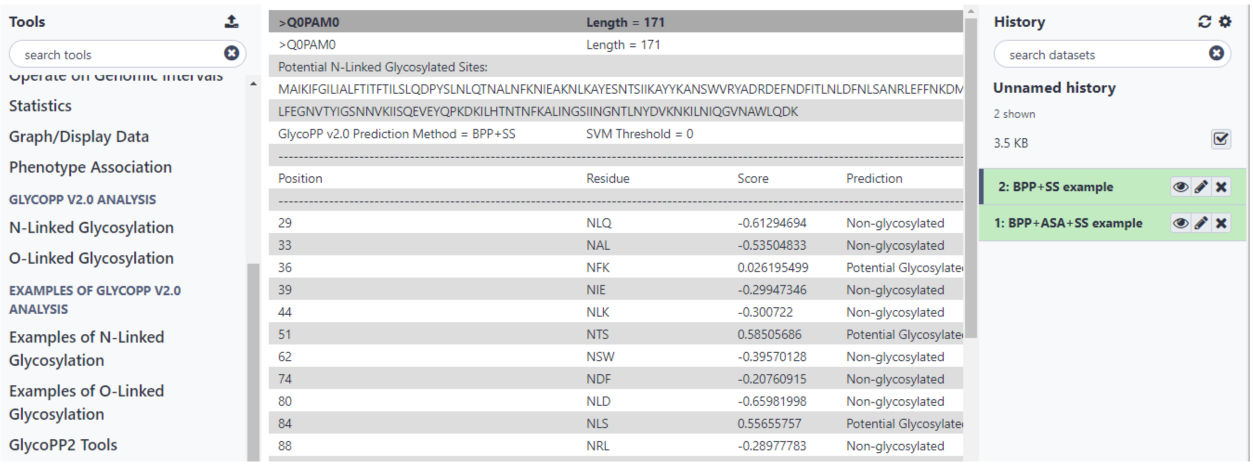
SVM Threshold value between -1.00 to +1.00. Default is 0

INFO: BPP+SS (Binary Profile Composition + Secondary Structure):
As sequence patterns of fixed length of 41-residues, we consider binary profile composition. Previous studies on eukaryotic glycoproteins suggested that the probability of finding glycosite was higher at positions where there was a secondary structure change. So, we applied the secondary structure along with DPC combination to get the higher result.

Please provide a fasta file for GlycoPP V2.0 BPP+SS prediction tool:-

This history is empty. You can **load your own data** or **get data from an external source**

Figure 16: *Example of GlycoPP V2.0 Analsis overview*

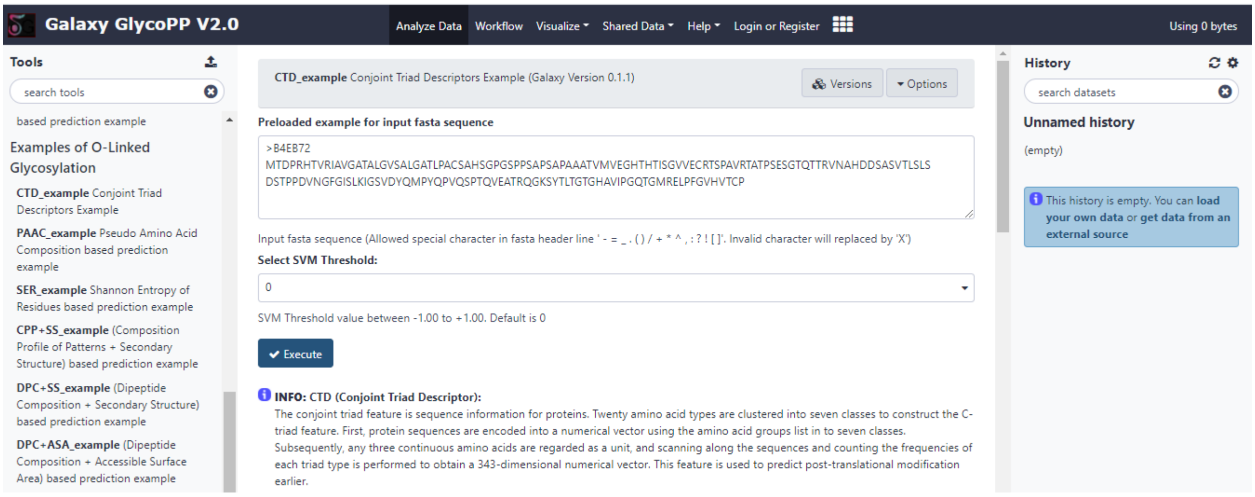


```
>Q0PAM0                                   Length = 171
>Q0PAM0                                   Length = 171
Potential N-Linked Glycosylated Sites:
MAIKIFGILIALFTITFTILSLQDPYSLNLQTNALNFKNIEAKNLKAYESNTSIIKAYYKANSWVRYADRDEFNDFITLNLDFNLSANRLEFFNKDM
LFEGNVTYIGSNNVKIISQEVEYQPKDKILHTNTNFKALINGSIINGNTLNYDVKNKILNIQGVNAWLQDK
GlycoPP v2.0 Prediction Method = BPP+SS   SVM Threshold = 0
----------------------------------------------------------------------------------------------------
Position                                  Residue        Score          Prediction
----------------------------------------------------------------------------------------------------
29                                        NLQ            -0.61294694    Non-glycosylated
33                                        NAL            -0.53504833    Non-glycosylated
36                                        NFK            0.026195499    Potential Glycosylate
39                                        NIE            -0.29947346    Non-glycosylated
44                                        NLK            -0.300722      Non-glycosylated
51                                        NTS            0.58505686     Potential Glycosylate
62                                        NSW            -0.39570128    Non-glycosylated
74                                        NDF            -0.20760915    Non-glycosylated
80                                        NLD            -0.65981998    Non-glycosylated
84                                        NLS            0.55655757     Potential Glycosylate
88                                        NRL            -0.28977783    Non-glycosylated
```

Figure 17: Example of GlycoPP V2.0 Analsis output for BPP+ASA+SS example (Binary Profile Composition+ Accessible Surface Area + Secondary Structure) forN-linked glycosylation

b) Example of O-linked Glycosylation

CTD_example Conjoint Triad Descriptors Example


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 0 bytes
Tools
search tools
based prediction example
Examples of O-Linked Glycosylation
CTD_example Conjoint Triad Descriptors Example
PAAC_example Pseudo Amino Acid Composition based prediction example
SER_example Shannon Entropy of Residues based prediction example
CPP+SS_example (Composition Profile of Patterns + Secondary Structure) based prediction example
DPC+SS_example (Dipeptide Composition + Secondary Structure) based prediction example
DPC+ASA_example (Dipeptide Composition + Accessible Surface Area) based prediction example
Versions
Options
Execute
History
search datasets
Unnamed history
(empty)

CTD_example Conjoint Triad Descriptors Example (Galaxy Version 0.1.1)

Preloaded example for input fasta sequence

```
>B4EB72
MTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTISGVVECRTSPAVRTATPSESGTQTTRVNAHDDSASVTLSLS
DSTPPDVNGFGISLKIGSVDYQMPYQPVQSPTQVEATRQGKSYTLTGTGHAVIPGQTGMRELPFGVHVTCP
```

Input fasta sequence (Allowed special character in fasta header line ' - = _ . ( ) / + * ^ , : ? ! [ ]'. Invalid character will replaced by 'X')

Select SVM Threshold:

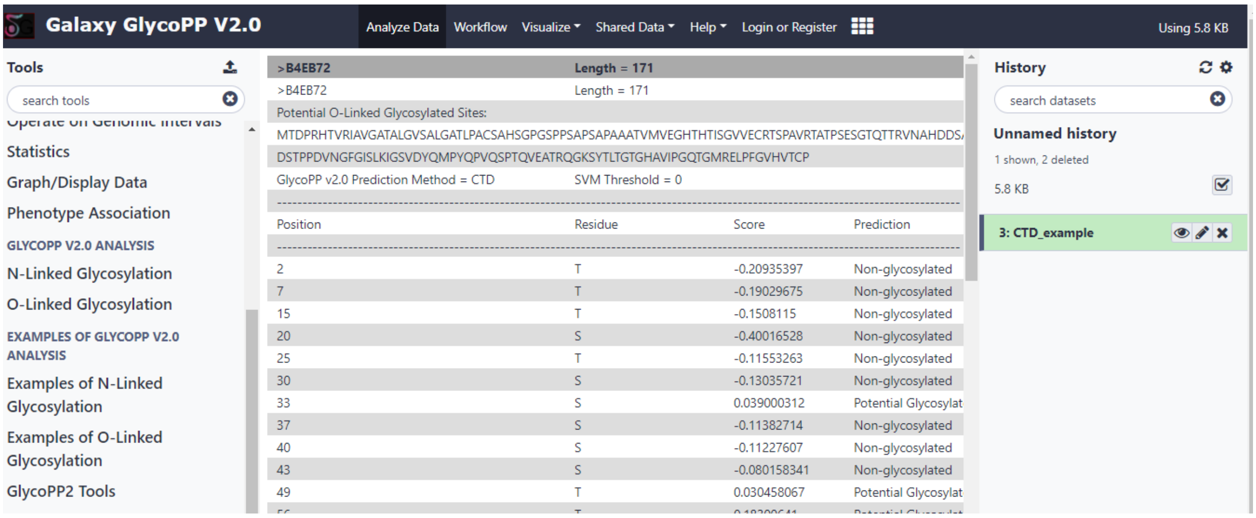
0

SVM Threshold value between -1.00 to +1.00. Default is 0

INFO: CTD (Conjoint Triad Descriptor):
The conjoint triad feature is sequence information for proteins. Twenty amino acid types are clustered into seven classes to construct the C-triad feature. First, protein sequences are encoded into a numerical vector using the amino acid groups list in to seven classes. Subsequently, any three continuous amino acids are regarded as a unit, and scanning along the sequences and counting the frequencies of each triad type is performed to obtain a 343-dimensional numerical vector. This feature is used to predict post-translational modification earlier.

This history is empty. You can load your own data or get data from an external source

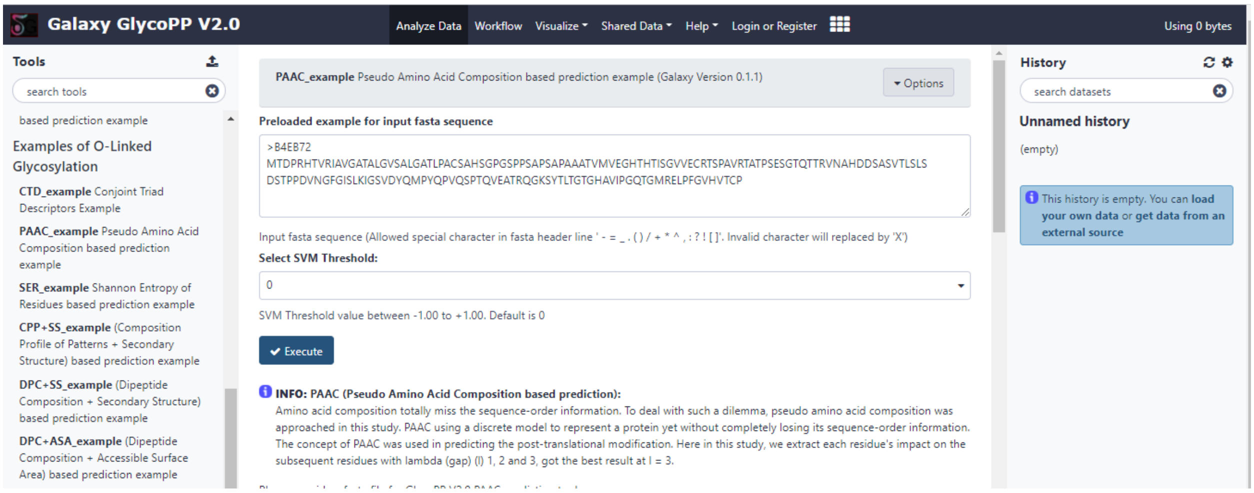
Figure 18: *Example of GlycoPP V2.0 Analsis overview*



```
>B4EB72                                       Length = 171
>B4EB72                                       Length = 171
Potential O-Linked Glycosylated Sites:
MTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTISGVVECRTSPAVRTATPSESGTQTTRVNAHDDS/
DSTPPDVNGFGISLKIGSVDYQMPYQPVQSPTQVEATRQGKSYTLTGTGHAVIPGQTGMRELPFGVHVTCP
GlycoPP v2.0 Prediction Method = CTD          SVM Threshold = 0
-----------------------------------------------------------------------------------------------------
Position                                      Residue          Score           Prediction
-----------------------------------------------------------------------------------------------------
2                                             T                -0.20935397     Non-glycosylated
7                                             T                -0.19029675     Non-glycosylated
15                                            T                -0.1508115      Non-glycosylated
20                                            S                -0.40016528     Non-glycosylated
25                                            T                -0.11553263     Non-glycosylated
30                                            S                -0.13035721     Non-glycosylated
33                                            S                0.039000312     Potential Glycosylat
37                                            S                -0.11382714     Non-glycosylated
40                                            S                -0.11227607     Non-glycosylated
43                                            S                -0.080158341    Non-glycosylated
49                                            T                0.030458067     Potential Glycosylat
```

Figure 19: Example of GlycoPP V2.0 Analsis output for CTD (Conjoint Triad Distributor) forO-linked glycosylation

PAAC_example Pseudo Amino Acid Composition based prediction example



based prediction example

Examples of O-Linked Glycosylation

CTD_example Conjoint Triad Descriptors Example

PAAC_example Pseudo Amino Acid Composition based prediction example

SER_example Shannon Entropy of Residues based prediction example

CPP+SS_example (Composition Profile of Patterns + Secondary Structure) based prediction example

DPC+SS_example (Dipeptide Composition + Secondary Structure) based prediction example

DPC+ASA_example (Dipeptide Composition + Accessible Surface Area) based prediction example

PAAC_example Pseudo Amino Acid Composition based prediction example (Galaxy Version 0.1.1)

Preloaded example for input fasta sequence

```
>B4EB72
MTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTISGVVECRTSPAVRTATPSESGTQTTRVNAHDDSASVTLSLS
DSTPPDVNGFGISLKIGSVDYQMPYQPVQSPTQVEATRQGKSYTLTGTGHAVIPGQTGMRELPFGVHVTCP
```

Input fasta sequence (Allowed special character in fasta header line ' - = _ . ( ) / + * ^ , : ? ! [ ]'. Invalid character will replaced by 'X')

Select SVM Threshold:

0

SVM Threshold value between -1.00 to +1.00. Default is 0

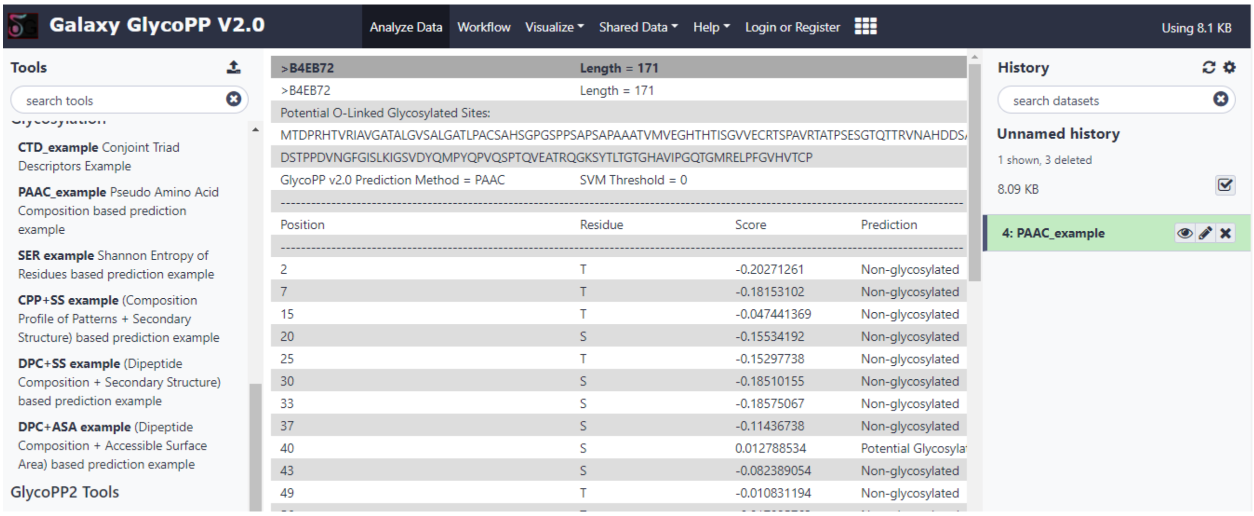
INFO: PAAC (Pseudo Amino Acid Composition based prediction):
Amino acid composition totally miss the sequence-order information. To deal with such a dilemma, pseudo amino acid composition was approached in this study. PAAC using a discrete model to represent a protein yet without completely losing its sequence-order information. The concept of PAAC was used in predicting the post-translational modification. Here in this study, we extract each residue's impact on the subsequent residues with lambda (gap) (l) 1, 2 and 3, got the best result at l = 3.

Unnamed history

(empty)

This history is empty. You can **load your own data** or **get data from an external source**

Figure 21: *Example of GlycoPP V2.0 Analsis overview*



CTD_example Conjoint Triad Descriptors Example

PAAC_example Pseudo Amino Acid Composition based prediction example

SER example Shannon Entropy of Residues based prediction example

CPP+SS example (Composition Profile of Patterns + Secondary Structure) based prediction example

DPC+SS example (Dipeptide Composition + Secondary Structure) based prediction example

DPC+ASA example (Dipeptide Composition + Accessible Surface Area) based prediction example

GlycoPP2 Tools

```
>B4EB72                                 Length = 171
>B4EB72                                 Length = 171
Potential O-Linked Glycosylated Sites:
MTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTISGVVECRTSPAVRTATPSESGTQTTRVNAHDDS
DSTPPDVNGFGISLKIGSVDYQMPYQPVQSPTQVEATRQGKSYTLTGTGHAVIPGQTGMRELPFGVHVTCP
GlycoPP v2.0 Prediction Method = PAAC    SVM Threshold = 0
-----------------------------------------------------------------------------------------------
Position                                Residue         Score           Prediction
-----------------------------------------------------------------------------------------------
2                                       T               -0.20271261     Non-glycosylated
7                                       T               -0.18153102     Non-glycosylated
15                                      T               -0.047441369    Non-glycosylated
20                                      S               -0.15534192     Non-glycosylated
25                                      T               -0.15297738     Non-glycosylated
30                                      S               -0.18510155     Non-glycosylated
33                                      S               -0.18575067     Non-glycosylated
37                                      S               -0.11436738     Non-glycosylated
40                                      S               0.012788534     Potential Glycosyla
43                                      S               -0.082389054    Non-glycosylated
49                                      T               -0.010831194    Non-glycosylated
```

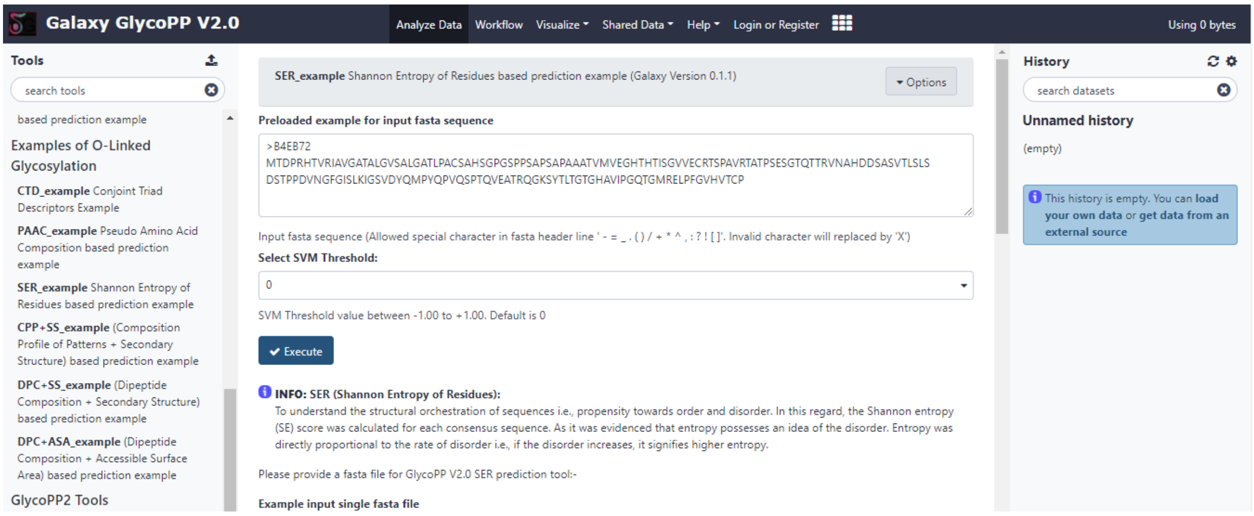
Unnamed history

1 shown, 3 deleted

8.09 KB

Figure 22: Example of GlycoPP V2.0 Analsis output for PAAC (Pseudo Amino Acid Composition) forO-linked glycosylation

SER example Shannon Entropy of Residues based prediction example


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 0 bytes
Tools
search tools
based prediction example
Examples of O-Linked Glycosylation
CTD_example Conjoint Triad Descriptors Example
PAAC_example Pseudo Amino Acid Composition based prediction example
SER_example Shannon Entropy of Residues based prediction example
CPP+SS_example (Composition Profile of Patterns + Secondary Structure) based prediction example
DPC+SS_example (Dipeptide Composition + Secondary Structure) based prediction example
DPC+ASA_example (Dipeptide Composition + Accessible Surface Area) based prediction example
GlycoPP2 Tools
Options
Execute
History
search datasets
Unnamed history
(empty)
This history is empty. You can load your own data or get data from an external source

SER_example Shannon Entropy of Residues based prediction example (Galaxy Version 0.1.1)

Preloaded example for input fasta sequence

```
>B4EB72
MTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTISGVVECRTSPAVRTATPSESGTQTTRVNAHDDSASVTLSLS
DSTPPDVNGFGISLKIGSVDYQMPYQPVQSPTQVEATRQGKSYTLTGTGHAVIPGQTGMRELPFGVHVTCP
```

Input fasta sequence (Allowed special character in fasta header line ' - = _ . ( ) / + * ^ , : ? ! [ ]'. Invalid character will replaced by 'X')

Select SVM Threshold:

0

SVM Threshold value between -1.00 to +1.00. Default is 0

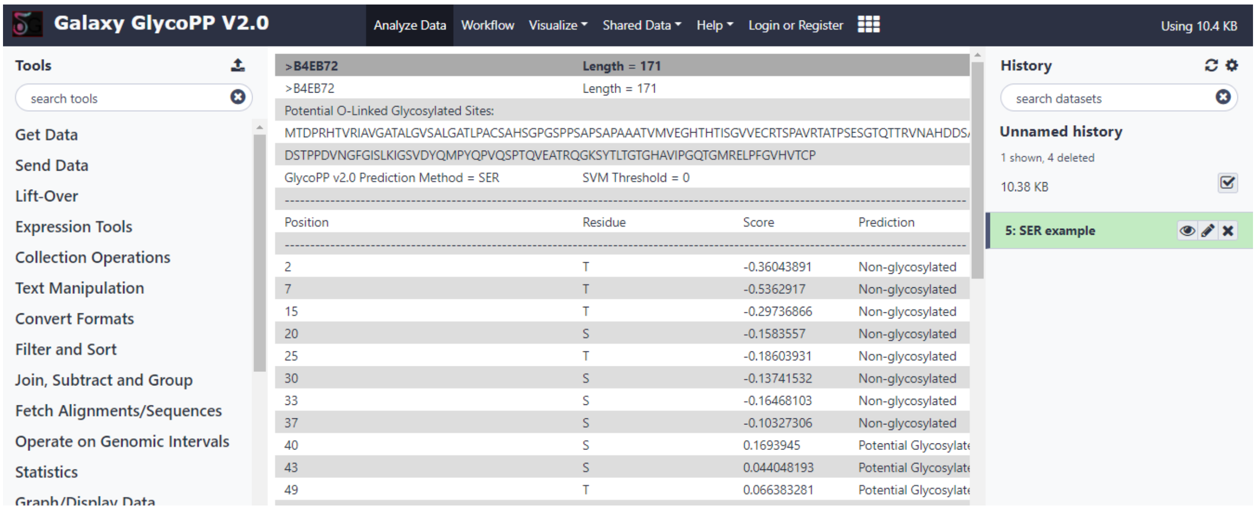
INFO: SER (Shannon Entropy of Residues):

To understand the structural orchestration of sequences i.e., propensity towards order and disorder. In this regard, the Shannon entropy (SE) score was calculated for each consensus sequence. As it was evidenced that entropy possesses an idea of the disorder. Entropy was directly proportional to the rate of disorder i.e., if the disorder increases, it signifies higher entropy.

Please provide a fasta file for GlycoPP V2.0 SER prediction tool:-

Example input single fasta file

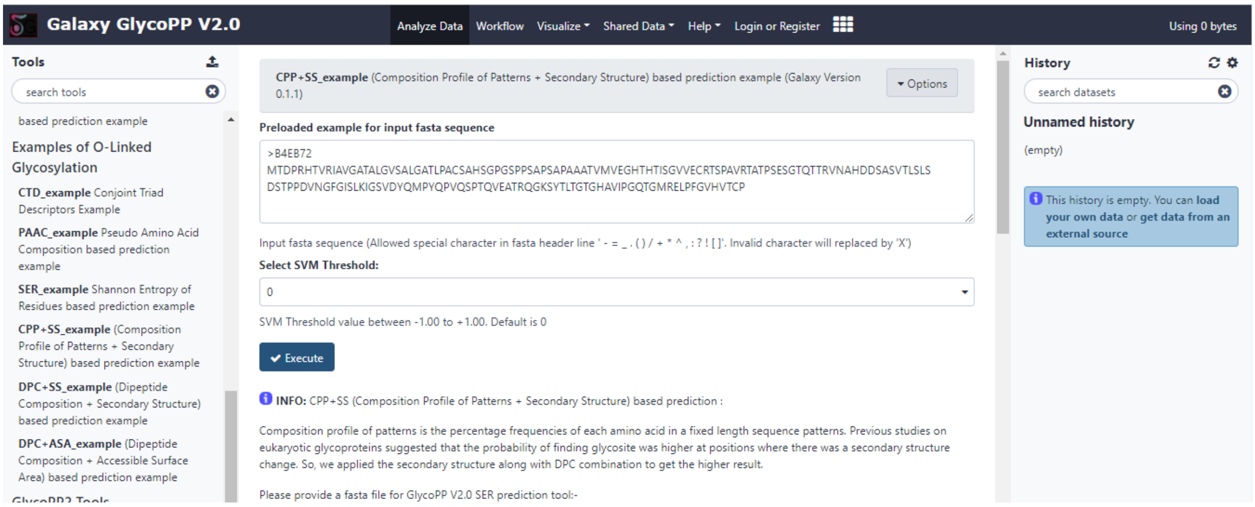
Figure 23: *Example of GlycoPP V2.0 Analsis overview*



```
>B4EB72                                  Length = 171
>B4EB72                                  Length = 171
Potential O-Linked Glycosylated Sites:
MTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTISGVVECRTSPAVRTATPSESGTQTTRVNAHDDS/
DSTPPDVNGFGISLKIGSVDYQMPYQPVQSPTQVEATRQGKSYTLTGTGHAVIPGQTGMRELPFGVHVTCP
GlycoPP v2.0 Prediction Method = SER     SVM Threshold = 0
-----------------------------------------------------------------------------------------------
Position                                 Residue     Score          Prediction
-----------------------------------------------------------------------------------------------
2                                        T           -0.36043891    Non-glycosylated
7                                        T           -0.5362917     Non-glycosylated
15                                       T           -0.29736866    Non-glycosylated
20                                       S           -0.1583557     Non-glycosylated
25                                       T           -0.18603931    Non-glycosylated
30                                       S           -0.13741532    Non-glycosylated
33                                       S           -0.16468103    Non-glycosylated
37                                       S           -0.10327306    Non-glycosylated
40                                       S           0.1693945      Potential Glycosylat
43                                       S           0.044048193    Potential Glycosylat
49                                       T           0.066383281    Potential Glycosylat
```

Figure 24: Example of GlycoPP V2.0 Analsis output for SER (Shannon Entropy of Residues) forO-linked glycosylation

CPP+SS example (Composition Profile of Patterns + Secondary Structure) based prediction example



based prediction example

Examples of O-Linked Glycosylation

CTD_example Conjoint Triad Descriptors Example

PAAC_example Pseudo Amino Acid Composition based prediction example

SER_example Shannon Entropy of Residues based prediction example

CPP+SS_example (Composition Profile of Patterns + Secondary Structure) based prediction example

DPC+SS_example (Dipeptide Composition + Secondary Structure) based prediction example

DPC+ASA_example (Dipeptide Composition + Accessible Surface Area) based prediction example

CPP+SS_example (Composition Profile of Patterns + Secondary Structure) based prediction example (Galaxy Version 0.1.1)

Preloaded example for input fasta sequence

```
>B4EB72
MTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTISGVVECRTSPAVRTATPSESGTQTTRVNAHDDSASVTLSLS
DSTPPDVNGFGISLKIGSVDYQMPYQPVQSPTQVEATRQGKSYTLTGTGHAVIPGQTGMRELPFGVHVTCP
```

Input fasta sequence (Allowed special character in fasta header line ' - = _ . ( ) / + * ^ , : ? ! [ ]'. Invalid character will replaced by 'X')

Select SVM Threshold:

0

SVM Threshold value between -1.00 to +1.00. Default is 0

INFO: CPP+SS (Composition Profile of Patterns + Secondary Structure) based prediction :

Composition profile of patterns is the percentage frequencies of each amino acid in a fixed length sequence patterns. Previous studies on eukaryotic glycoproteins suggested that the probability of finding glycosite was higher at positions where there was a secondary structure change. So, we applied the secondary structure along with DPC combination to get the higher result.

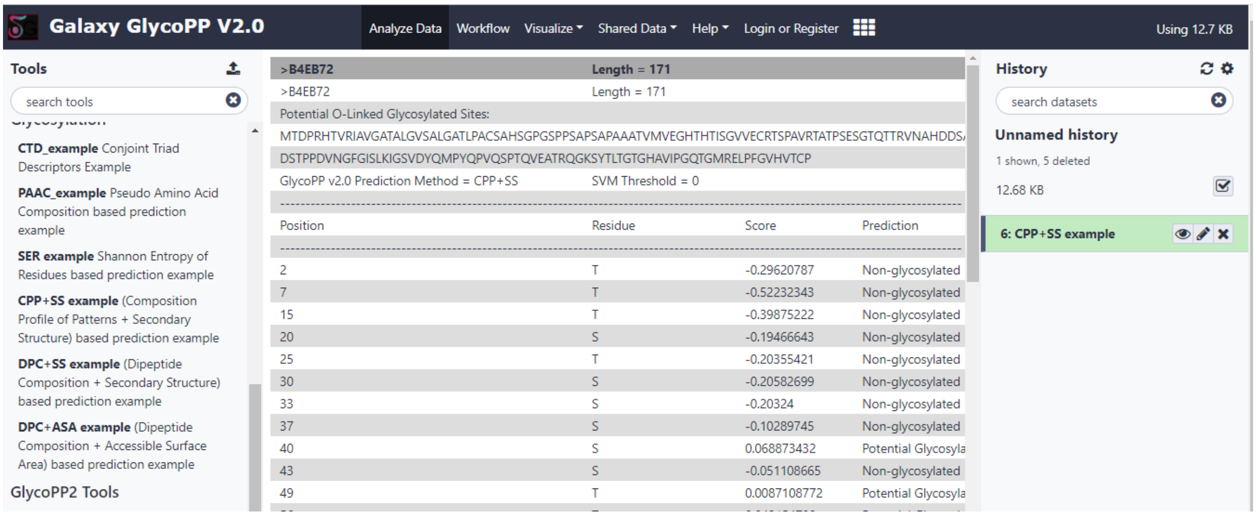
Please provide a fasta file for GlycoPP V2.0 SER prediction tool:-

Unnamed history

(empty)

This history is empty. You can **load your own data** or **get data from an external source**

Figure 25: Example of GlycoPP V2.0 Analsis overview



CTD_example Conjoint Triad Descriptors Example

PAAC_example Pseudo Amino Acid Composition based prediction example

SER example Shannon Entropy of Residues based prediction example

CPP+SS example (Composition Profile of Patterns + Secondary Structure) based prediction example

DPC+SS example (Dipeptide Composition + Secondary Structure) based prediction example

DPC+ASA example (Dipeptide Composition + Accessible Surface Area) based prediction example

GlycoPP2 Tools

```
>B4EB72                                      Length = 171
>B4EB72                                      Length = 171
Potential O-Linked Glycosylated Sites:
MTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTISGVVECRTSPAVRTATPSESGTQTTRVNAHDDS/
DSTPPDVNGFGISLKIGSVDYQMPYQPVQSPTQVEATRQGKSYTLTGTGHAVIPGQTGMRELPFGVHVTCP
GlycoPP v2.0 Prediction Method = CPP+SS       SVM Threshold = 0
------------------------------------------------------------------------------------------------
Position                                      Residue          Score           Prediction
------------------------------------------------------------------------------------------------
2                                             T                -0.29620787     Non-glycosylated
7                                             T                -0.52232343     Non-glycosylated
15                                            T                -0.39875222     Non-glycosylated
20                                            S                -0.19466643     Non-glycosylated
25                                            T                -0.20355421     Non-glycosylated
30                                            S                -0.20582699     Non-glycosylated
33                                            S                -0.20324        Non-glycosylated
37                                            S                -0.10289745     Non-glycosylated
40                                            S                0.068873432     Potential Glycosyla
43                                            S                -0.051108665    Non-glycosylated
49                                            T                0.0087108772    Potential Glycosyla
```

Unnamed history

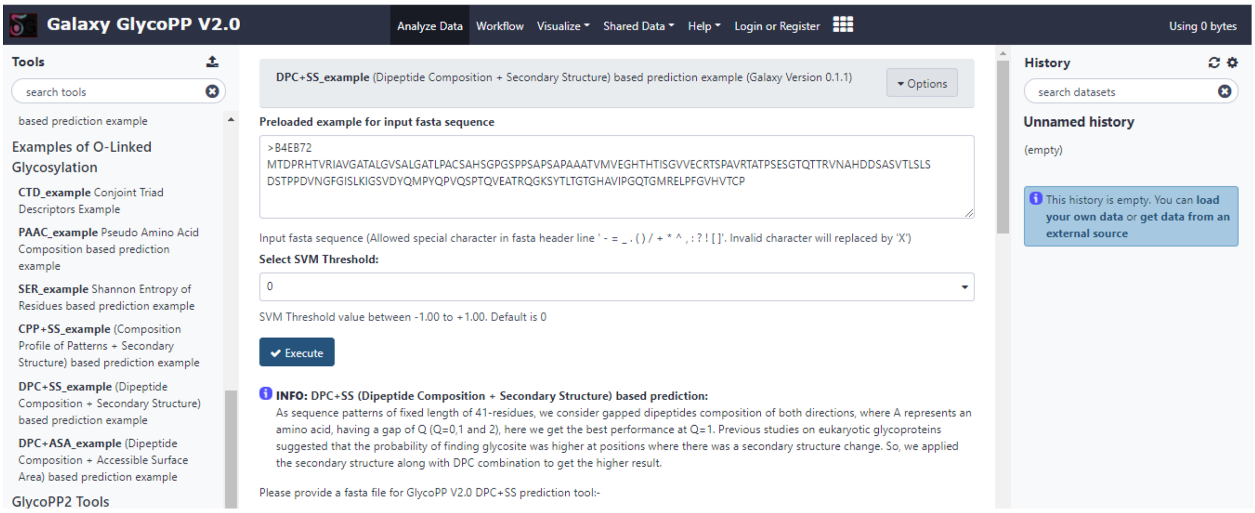
1 shown, 5 deleted

12.68 KB

6: CPP+SS example

Figure 26: Example of GlycoPP V2.0 Analsis output for CPP+SS (Composition Profile of Patterns + Secondary Structure) forO-linked glycosylation

DPC+SS example (Dipeptide Composition + Secondary Structure)based prediction example


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 0 bytes
Tools
search tools
based prediction example
Examples of O-Linked Glycosylation
CTD_example Conjoint Triad Descriptors Example
PAAC_example Pseudo Amino Acid Composition based prediction example
SER_example Shannon Entropy of Residues based prediction example
CPP+SS_example (Composition Profile of Patterns + Secondary Structure) based prediction example
DPC+SS_example (Dipeptide Composition + Secondary Structure) based prediction example
DPC+ASA_example (Dipeptide Composition + Accessible Surface Area) based prediction example
GlycoPP2 Tools
History
search datasets
Unnamed history
(empty)
This history is empty. You can load your own data or get data from an external source
Options
Execute

DPC+SS_example (Dipeptide Composition + Secondary Structure) based prediction example (Galaxy Version 0.1.1)

Preloaded example for input fasta sequence

```
>B4EB72
MTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTISGVVECRTSPAVRTATPSESGTQTTRVNAHDDSASVTLSLS
DSTPPDVNGFGISLKIGSVDYQMPYQPVQSPTQVEATRQGKSYTLTGTGHAVIPGQTGMRELPFGVHVTCP
```

Input fasta sequence (Allowed special character in fasta header line ' - = _ . ( ) / + * ^ , : ? ! [ ]'. Invalid character will replaced by 'X')

Select SVM Threshold:

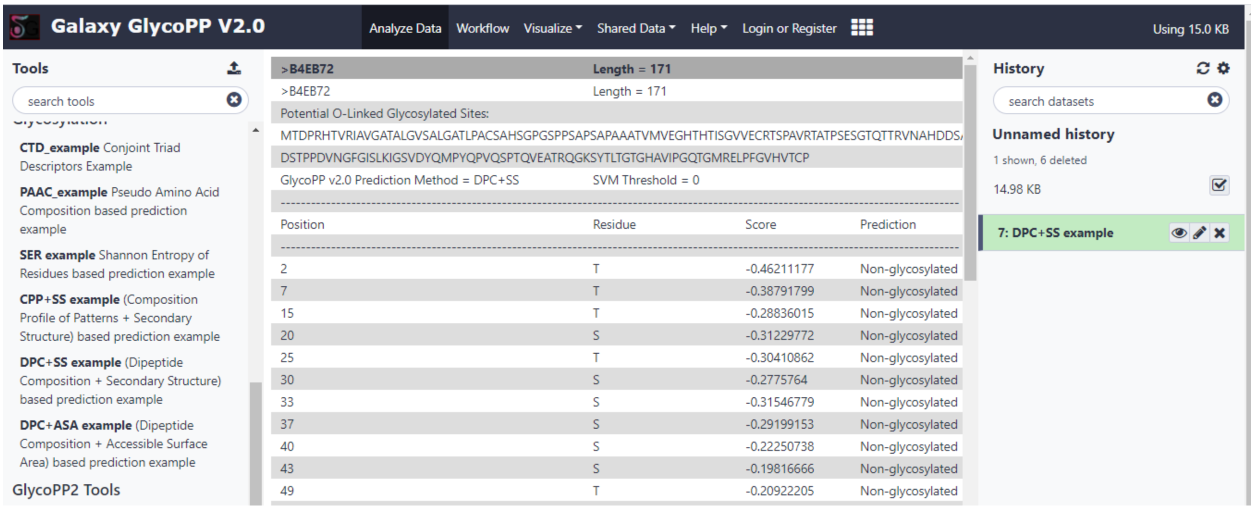
0

SVM Threshold value between -1.00 to +1.00. Default is 0

INFO: DPC+SS (Dipeptide Composition + Secondary Structure) based prediction:
As sequence patterns of fixed length of 41-residues, we consider gapped dipeptides composition of both directions, where A represents an amino acid, having a gap of Q (Q=0,1 and 2), here we get the best performance at Q=1. Previous studies on eukaryotic glycoproteins suggested that the probability of finding glycosite was higher at positions where there was a secondary structure change. So, we applied the secondary structure along with DPC combination to get the higher result.

Please provide a fasta file for GlycoPP V2.0 DPC+SS prediction tool:-

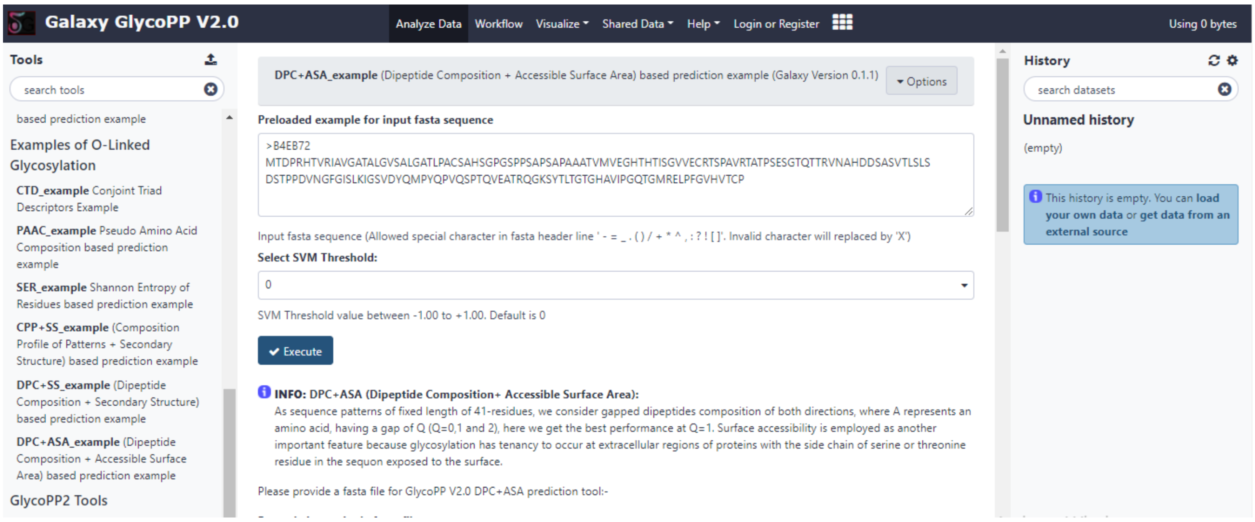
Figure 27: *Example of GlycoPP V2.0 Analsis overview*


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 15.0 KB
Tools
search tools
CTD_example Conjoint Triad Descriptors Example
PAAC_example Pseudo Amino Acid Composition based prediction example
SER example Shannon Entropy of Residues based prediction example
CPP+SS example (Composition Profile of Patterns + Secondary Structure) based prediction example
DPC+SS example (Dipeptide Composition + Secondary Structure) based prediction example
DPC+ASA example (Dipeptide Composition + Accessible Surface Area) based prediction example
GlycoPP2 Tools
History
search datasets
Unnamed history
1 shown, 6 deleted
14.98 KB
7: DPC+SS example

```
>B4EB72                                      Length = 171
>B4EB72                                      Length = 171
Potential O-Linked Glycosylated Sites:
MTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTISGVVECRTSPAVRTATPSESGTQTTRVNAHDDS/
DSTPPDVNGFGISLKIGSVDYQMPYQPVQSPTQVEATRQGKSYTLTGTGHAVIPGQTGMRELPFGVHVTCP
GlycoPP v2.0 Prediction Method = DPC+SS      SVM Threshold = 0
-------------------------------------------------------------------------------------------
Position                                     Residue         Score          Prediction
-------------------------------------------------------------------------------------------
2                                            T               -0.46211177    Non-glycosylated
7                                            T               -0.38791799    Non-glycosylated
15                                           T               -0.28836015    Non-glycosylated
20                                           S               -0.31229772    Non-glycosylated
25                                           T               -0.30410862    Non-glycosylated
30                                           S               -0.2775764     Non-glycosylated
33                                           S               -0.31546779    Non-glycosylated
37                                           S               -0.29199153    Non-glycosylated
40                                           S               -0.22250738    Non-glycosylated
43                                           S               -0.19816666    Non-glycosylated
49                                           T               -0.20922205    Non-glycosylated
```

Figure 28: Example of GlycoPP V2.0 Analsis output for DPC+SS(Dipeptide Composition + Secondary Structure)forO-linked glycosylation

DPC+ASA example (Dipeptide Composition + Accesible surface Area)based prediction example


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 0 bytes
Tools
search tools
based prediction example
Examples of O-Linked Glycosylation
CTD_example Conjoint Triad Descriptors Example
PAAC_example Pseudo Amino Acid Composition based prediction example
SER_example Shannon Entropy of Residues based prediction example
CPP+SS_example (Composition Profile of Patterns + Secondary Structure) based prediction example
DPC+SS_example (Dipeptide Composition + Secondary Structure) based prediction example
DPC+ASA_example (Dipeptide Composition + Accessible Surface Area) based prediction example
GlycoPP2 Tools
History
search datasets
Unnamed history
(empty)
This history is empty. You can load your own data or get data from an external source
Options
Execute

DPC+ASA_example (Dipeptide Composition + Accessible Surface Area) based prediction example (Galaxy Version 0.1.1)

Preloaded example for input fasta sequence

```
>B4EB72
MTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTISGVVECRTSPAVRTATPSESGTQTTRVNAHDDSASVTLSLS
DSTPPDVNGFGISLKIGSVDYQMPYQPVQSPTQVEATRQGKSYTLTGTGHAVIPGQTGMRELPFGVHVTCP
```

Input fasta sequence (Allowed special character in fasta header line ' - = _ . ( ) / + * ^ , : ? ! [ ]'. Invalid character will replaced by 'X')

Select SVM Threshold:

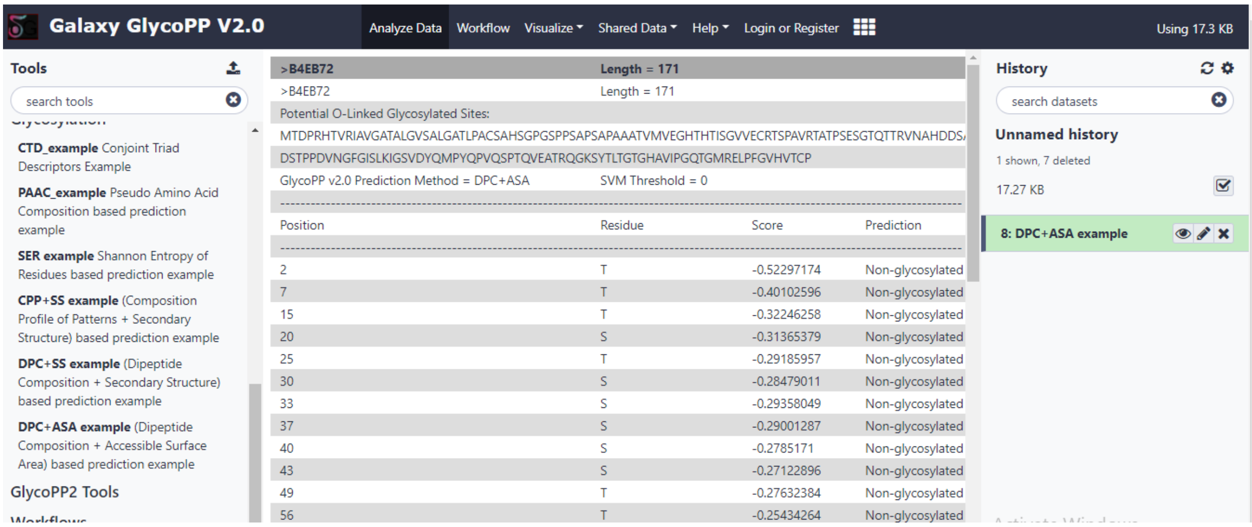
0

SVM Threshold value between -1.00 to +1.00. Default is 0

INFO: **DPC+ASA (Dipeptide Composition+ Accessible Surface Area):**
As sequence patterns of fixed length of 41-residues, we consider gapped dipeptides composition of both directions, where A represents an amino acid, having a gap of Q (Q=0,1 and 2), here we get the best performance at Q=1. Surface accessibility is employed as another important feature because glycosylation has tenancy to occur at extracellular regions of proteins with the side chain of serine or threonine residue in the sequon exposed to the surface.

Please provide a fasta file for GlycoPP V2.0 DPC+ASA prediction tool:-

Figure 29: *Example of GlycoPP V2.0 Analsis overview*


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 17.3 KB
Tools
search tools
CTD_example Conjoint Triad Descriptors Example
PAAC_example Pseudo Amino Acid Composition based prediction example
SER example Shannon Entropy of Residues based prediction example
CPP+SS example (Composition Profile of Patterns + Secondary Structure) based prediction example
DPC+SS example (Dipeptide Composition + Secondary Structure) based prediction example
DPC+ASA example (Dipeptide Composition + Accessible Surface Area) based prediction example
GlycoPP2 Tools
History
search datasets
Unnamed history
1 shown, 7 deleted
17.27 KB
8: DPC+ASA example

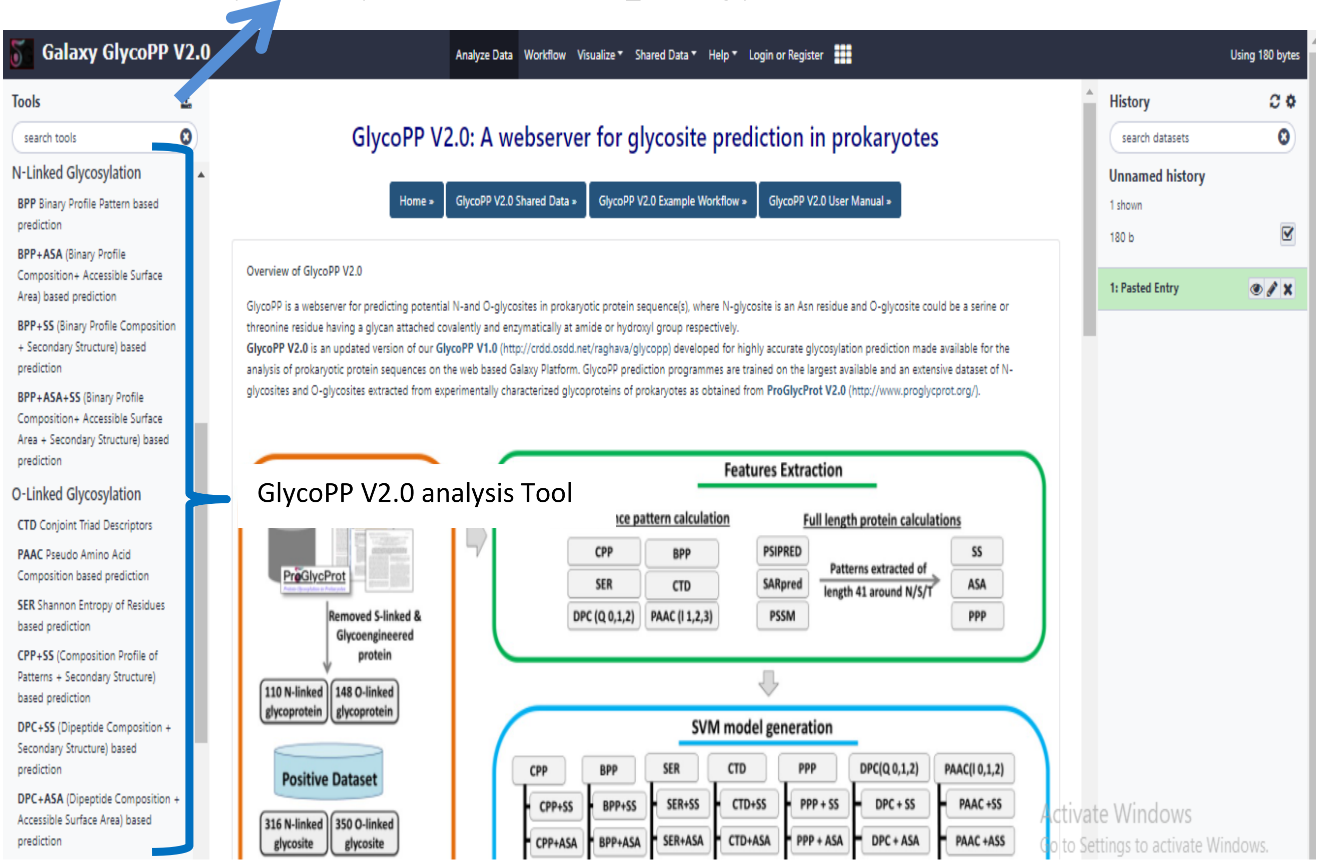
```
>B4EB72                                  Length = 171
>B4EB72                                  Length = 171
Potential O-Linked Glycosylated Sites:
MTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTISGVVECRTSPAVRTATPSESGTQTTRVNAHDDS/
DSTPPDVNGFGISLKIGSVDYQMPYQPVQSPTQVEATRQGKSYTLTGTGHAVIPGQTGMRELPFGVHVTCP
GlycoPP v2.0 Prediction Method = DPC+ASA  SVM Threshold = 0
------------------------------------------------------------------------------------------
Position                                 Residue          Score          Prediction
------------------------------------------------------------------------------------------
2                                        T                -0.52297174    Non-glycosylated
7                                        T                -0.40102596    Non-glycosylated
15                                       T                -0.32246258    Non-glycosylated
20                                       S                -0.31365379    Non-glycosylated
25                                       T                -0.29185957    Non-glycosylated
30                                       S                -0.28479011    Non-glycosylated
33                                       S                -0.29358049    Non-glycosylated
37                                       S                -0.29001287    Non-glycosylated
40                                       S                -0.2785171     Non-glycosylated
43                                       S                -0.27122896    Non-glycosylated
49                                       T                -0.27632384    Non-glycosylated
56                                       T                -0.25434264    Non-glycosylated
```

Figure 30: Example of GlycoPP V2.0 Analsis output for DPC+ASA example (Dipeptide Composition + Accesible surface Area) forO-linked glycosylation

2. Glycopp V2.0 Analysis

A user can upload the fasta file using the Galaxy tool **Get Data as we shown in figure 8** this fasta files can be used for GlycoPP V2.0 Analysis.The following figure shows the file upload method.

Get Data Galaxy tool for uploading the fasta protein sequence for prediction of N- and O_linked glycosite



Galaxy GlycoPP V2.0

N-Linked Glycosylation

- **BPP** Binary Profile Pattern based prediction
- **BPP+ASA** (Binary Profile Composition+ Accessible Surface Area) based prediction
- **BPP+SS** (Binary Profile Composition + Secondary Structure) based prediction
- **BPP+ASA+SS** (Binary Profile Composition+ Accessible Surface Area + Secondary Structure) based prediction

O-Linked Glycosylation

- **CTD** Conjoint Triad Descriptors
- **PAAC** Pseudo Amino Acid Composition based prediction
- **SER** Shannon Entropy of Residues based prediction
- **CPP+SS** (Composition Profile of Patterns + Secondary Structure) based prediction
- **DPC+SS** (Dipeptide Composition + Secondary Structure) based prediction
- **DPC+ASA** (Dipeptide Composition + Accessible Surface Area) based prediction

GlycoPP V2.0: A webserver for glycosite prediction in prokaryotes

Home »
GlycoPP V2.0 Shared Data »
GlycoPP V2.0 Example Workflow »
GlycoPP V2.0 User Manual »

Overview of GlycoPP V2.0

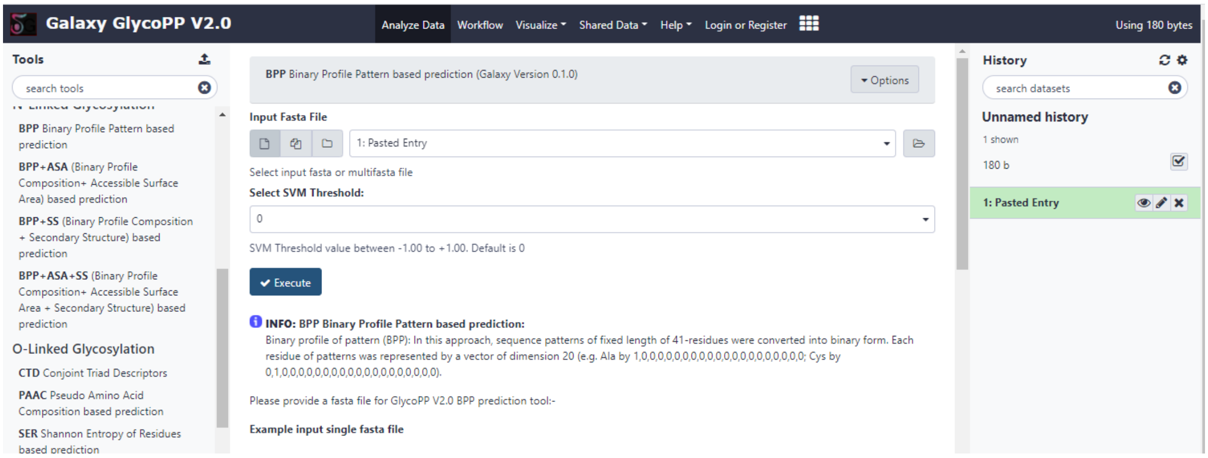
GlycoPP is a webserver for predicting potential N-and O-glycosites in prokaryotic protein sequence(s), where N-glycosite is an Asn residue and O-glycosite could be a serine or threonine residue having a glycan attached covalently and enzymatically at amide or hydroxyl group respectively.
GlycoPP V2.0 is an updated version of our **GlycoPP V1.0** (http://crdd.osdd.net/raghava/glycopp) developed for highly accurate glycosylation prediction made available for the analysis of prokaryotic protein sequences on the web based Galaxy Platform. GlycoPP prediction programmes are trained on the largest available and an extensive dataset of N-glycosites and O-glycosites extracted from experimentally characterized glycoproteins of prokaryotes as obtained from **ProGlycProt V2.0** (http://www.proglycprot.org/).

GlycoPP V2.0 analysis Tool

Figure31 : GlycoPP V2.0 Analsis overview

a) N-linked Glycosylation

BPP Binary Profile Pattern based prediction


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 180 bytes
Tools
search tools
BPP Binary Profile Pattern based prediction
BPP+ASA (Binary Profile Composition+ Accessible Surface Area) based prediction
BPP+SS (Binary Profile Composition + Secondary Structure) based prediction
BPP+ASA+SS (Binary Profile Composition+ Accessible Surface Area + Secondary Structure) based prediction
O-Linked Glycosylation
CTD Conjoint Triad Descriptors
PAAC Pseudo Amino Acid Composition based prediction
SER Shannon Entropy of Residues based prediction
History
search datasets
Unnamed history
1 shown
180 b
1: Pasted Entry
Options

BPP Binary Profile Pattern based prediction (Galaxy Version 0.1.0)

Input Fasta File

1: Pasted Entry

Select input fasta or multifasta file

Select SVM Threshold:

0

SVM Threshold value between -1.00 to +1.00. Default is 0

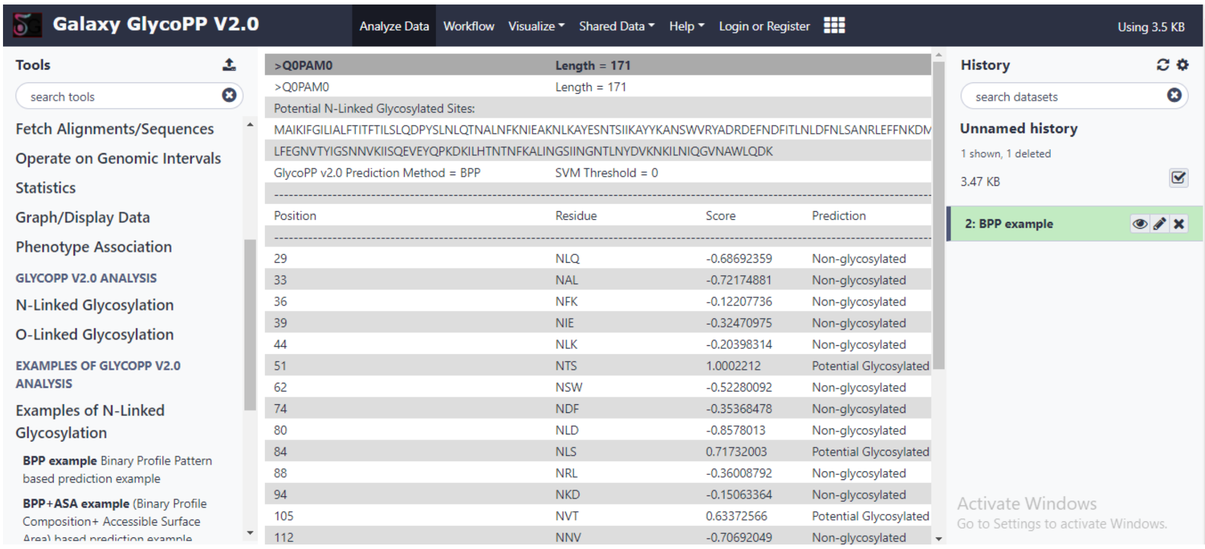
✔ Execute

INFO: BPP Binary Profile Pattern based prediction:
Binary profile of pattern (BPP): In this approach, sequence patterns of fixed length of 41-residues were converted into binary form. Each residue of patterns was represented by a vector of dimension 20 (e.g. Ala by 1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0; Cys by 0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0).

Please provide a fasta file for GlycoPP V2.0 BPP prediction tool:-

Example input single fasta file

Figure-32 : *GlycoPP V2.0 Analsis overview for BPP*



>Q0PAM0 **Length = 171**

>Q0PAM0 Length = 171

Potential N-Linked Glycosylated Sites:

MAIKIFGILIALFTITFTILSLQDPYSLNLQTNALNFKNIEAKNLKAYESNTSIIKAYYKANSWVRYADRDEFNDFITLNLDFNLSANRLEFFNKDM

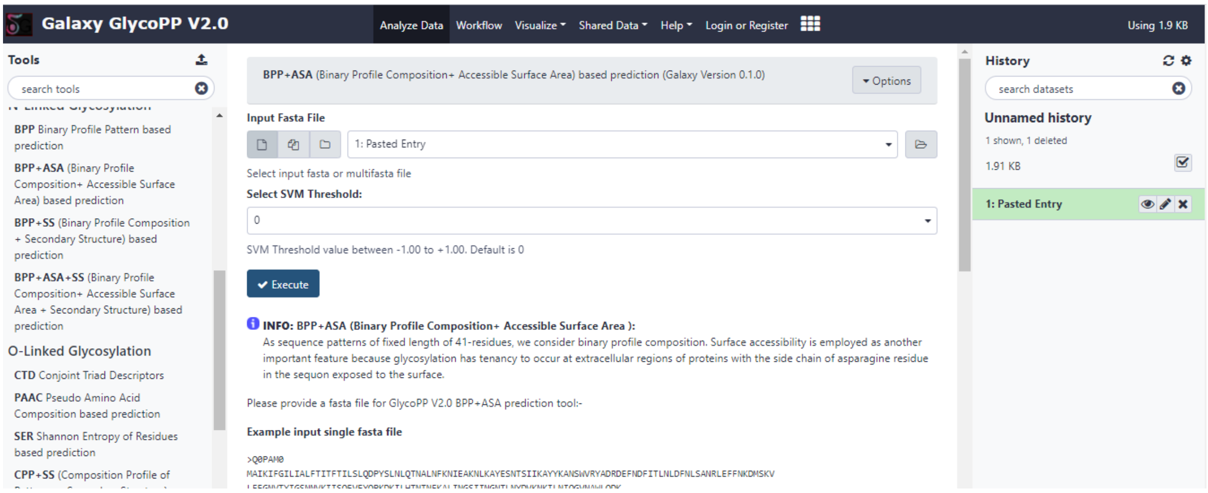
LFEGNVTYIGSNNVKIISQEVEYQPKDKILHTNTNFKALINGSIINGNTLNYDVKNKILNIQGVNAWLQDK

GlycoPP v2.0 Prediction Method = BPP SVM Threshold = 0

Position	Residue	Score	Prediction
29	NLQ	-0.68692359	Non-glycosylated
33	NAL	-0.72174881	Non-glycosylated
36	NFK	-0.12207736	Non-glycosylated
39	NIE	-0.32470975	Non-glycosylated
44	NLK	-0.20398314	Non-glycosylated
51	NTS	1.0002212	Potential Glycosylated
62	NSW	-0.52280092	Non-glycosylated
74	NDF	-0.35368478	Non-glycosylated
80	NLD	-0.8578013	Non-glycosylated
84	NLS	0.71732003	Potential Glycosylated
88	NRL	-0.36008792	Non-glycosylated
94	NKD	-0.15063364	Non-glycosylated
105	NVT	0.63372566	Potential Glycosylated
112	NNV	-0.70692049	Non-glycosylated

Figure33: *GlycoPP V2.0 Analsis output for BPP (Binary Profile Pattern) forN-linked glycosylation*

BPP+ASA (Binary Profile Composition+ Accessible Surface Area) based prediction


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 1.9 KB
Tools
search tools
BPP Binary Profile Pattern based prediction
BPP+ASA (Binary Profile Composition+ Accessible Surface Area) based prediction
BPP+SS (Binary Profile Composition + Secondary Structure) based prediction
BPP+ASA+SS (Binary Profile Composition+ Accessible Surface Area + Secondary Structure) based prediction
O-Linked Glycosylation
CTD Conjoint Triad Descriptors
PAAC Pseudo Amino Acid Composition based prediction
SER Shannon Entropy of Residues based prediction
CPP+SS (Composition Profile of
History
search datasets
Unnamed history
1 shown, 1 deleted
1.91 KB
1: Pasted Entry
Options
Execute

BPP+ASA (Binary Profile Composition+ Accessible Surface Area) based prediction (Galaxy Version 0.1.0)

Input Fasta File

1: Pasted Entry

Select input fasta or multifasta file

Select SVM Threshold:

0

SVM Threshold value between -1.00 to +1.00. Default is 0

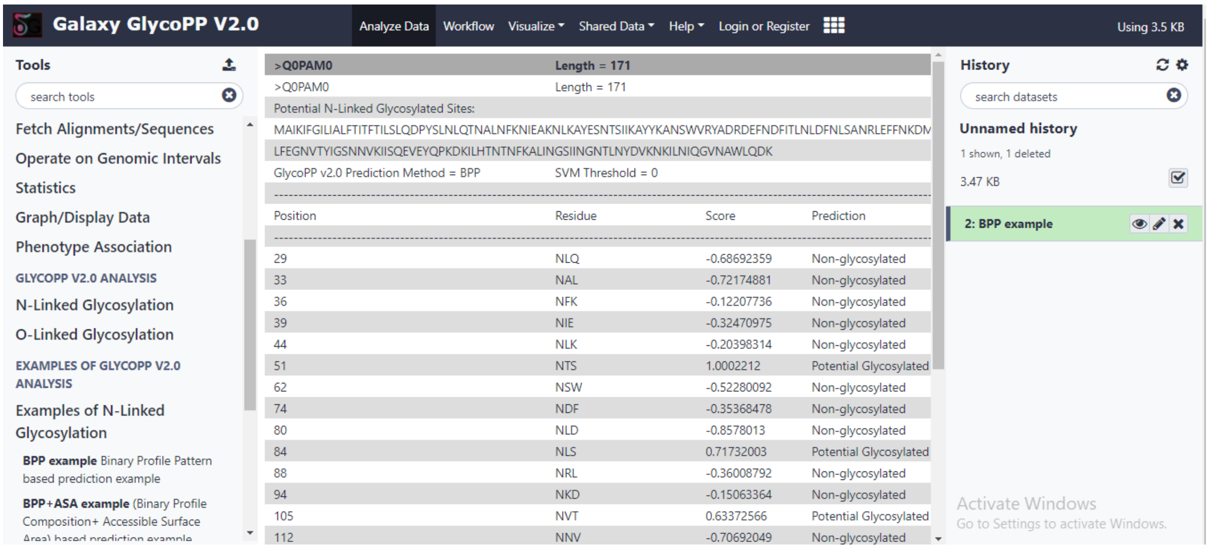
INFO: BPP+ASA (Binary Profile Composition+ Accessible Surface Area):
As sequence patterns of fixed length of 41-residues, we consider binary profile composition. Surface accessibility is employed as another important feature because glycosylation has tenancy to occur at extracellular regions of proteins with the side chain of asparagine residue in the sequon exposed to the surface.

Please provide a fasta file for GlycoPP V2.0 BPP+ASA prediction tool:-

Example input single fasta file

```
>Q0PAM0
MAIKIFGILIALFTITFTILSLQDPYSLNLQTNALNFKNIEAKNLKAYESNTSIIKAYYKANSWVRYADRDEFNDFITLNLDFNLSANRLEFFNKDMSKV
```

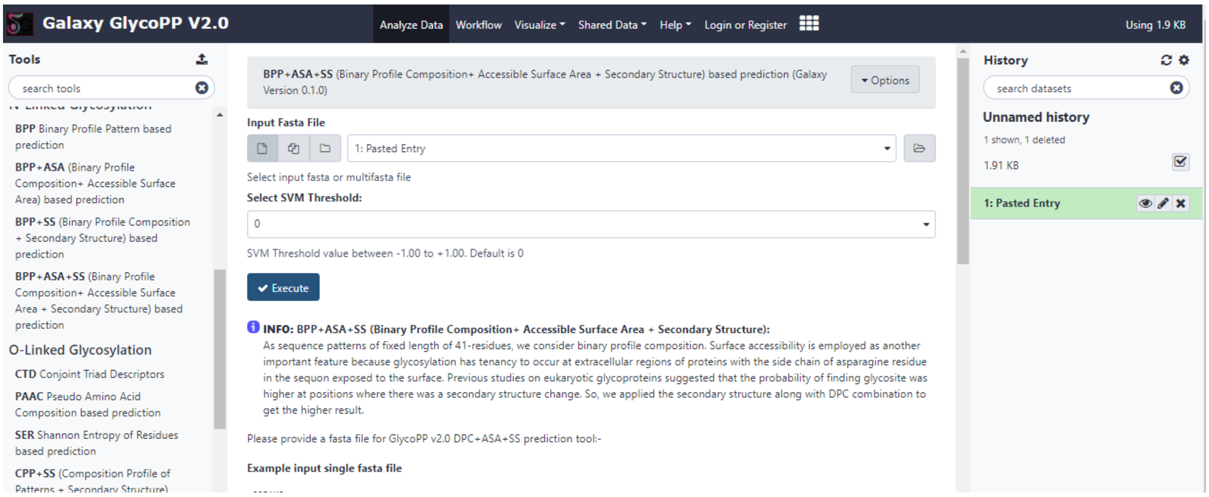
Figure 34: *GlycoPP V2.0 Analsis*



```
>Q0PAM0                                   Length = 171
>Q0PAM0                                   Length = 171
Potential N-Linked Glycosylated Sites:
MAIKIFGILIALFTITFTILSLQDPYSLNLQTNALNFKNIEAKNLKAYESNTSIIKAYYKANSWVRYADRDEFNDFITLNLDFNLSANRLEFFNKDM
LFEGNVTYIGSNNVKIISQEVEYQPKDKILHTNTNFKALINGSIINGNTLNYDVKNKILNIQGVNAWLQDK
GlycoPP v2.0 Prediction Method = BPP      SVM Threshold = 0
----------------------------------------------------------------------------------------
Position                                  Residue          Score          Prediction
----------------------------------------------------------------------------------------
29                                        NLQ              -0.68692359    Non-glycosylated
33                                        NAL              -0.72174881    Non-glycosylated
36                                        NFK              -0.12207736    Non-glycosylated
39                                        NIE              -0.32470975    Non-glycosylated
44                                        NLK              -0.20398314    Non-glycosylated
51                                        NTS              1.0002212      Potential Glycosylated
62                                        NSW              -0.52280092    Non-glycosylated
74                                        NDF              -0.35368478    Non-glycosylated
80                                        NLD              -0.8578013     Non-glycosylated
84                                        NLS              0.71732003     Potential Glycosylated
88                                        NRL              -0.36008792    Non-glycosylated
94                                        NKD              -0.15063364    Non-glycosylated
105                                       NVT              0.63372566     Potential Glycosylated
112                                       NNV              -0.70692049    Non-glycosylated
```

Figure 35: GlycoPP V2.0 Analsis output for BPP+ASA (Binary Profile Pattern+ Accessible Surface Are) forN-linked glycosylation

BPP+ASA+SS (Binary Profile Composition+ Accessible Surface Area + Secondary Structure) based prediction


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 1.9 KB
Tools
search tools
BPP Binary Profile Pattern based prediction
BPP+ASA (Binary Profile Composition+ Accessible Surface Area) based prediction
BPP+SS (Binary Profile Composition + Secondary Structure) based prediction
BPP+ASA+SS (Binary Profile Composition+ Accessible Surface Area + Secondary Structure) based prediction
O-Linked Glycosylation
CTD Conjoint Triad Descriptors
PAAC Pseudo Amino Acid Composition based prediction
SER Shannon Entropy of Residues based prediction
CPP+SS (Composition Profile of Patterns + Secondary Structure)
Options
Execute
History
search datasets
Unnamed history
1 shown, 1 deleted
1.91 KB
1: Pasted Entry

BPP+ASA+SS (Binary Profile Composition+ Accessible Surface Area + Secondary Structure) based prediction (Galaxy Version 0.1.0)

Input Fasta File

1: Pasted Entry

Select input fasta or multifasta file

Select SVM Threshold:

0

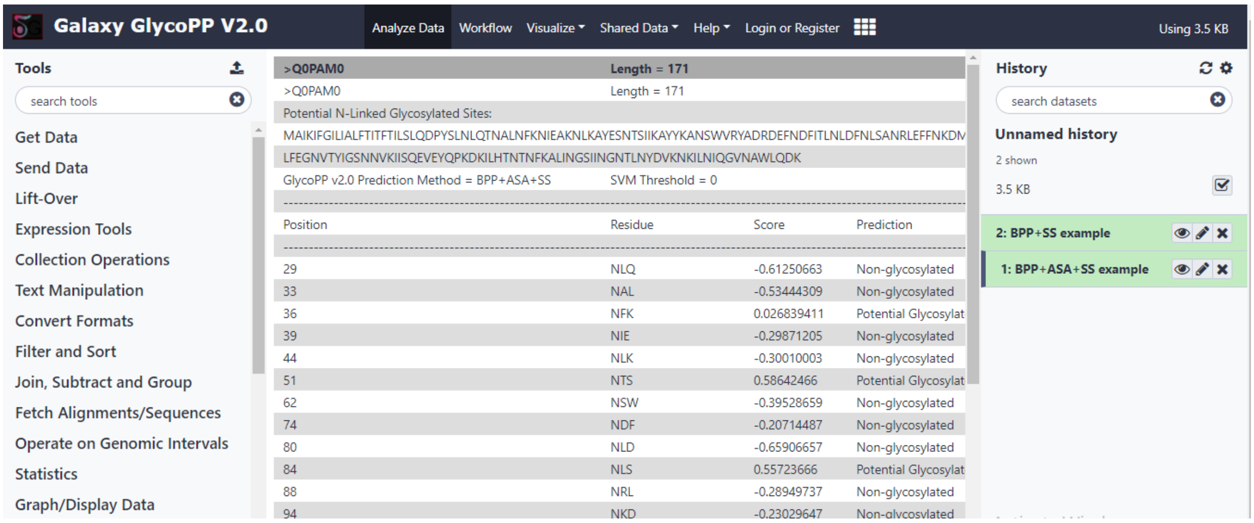
SVM Threshold value between -1.00 to +1.00. Default is 0

INFO: BPP+ASA+SS (Binary Profile Composition+ Accessible Surface Area + Secondary Structure):
As sequence patterns of fixed length of 41-residues, we consider binary profile composition. Surface accessibility is employed as another important feature because glycosylation has tenancy to occur at extracellular regions of proteins with the side chain of asparagine residue in the sequon exposed to the surface. Previous studies on eukaryotic glycoproteins suggested that the probability of finding glycosite was higher at positions where there was a secondary structure change. So, we applied the secondary structure along with DPC combination to get the higher result.

Please provide a fasta file for GlycoPP v2.0 DPC+ASA+SS prediction tool:-

Example input single fasta file

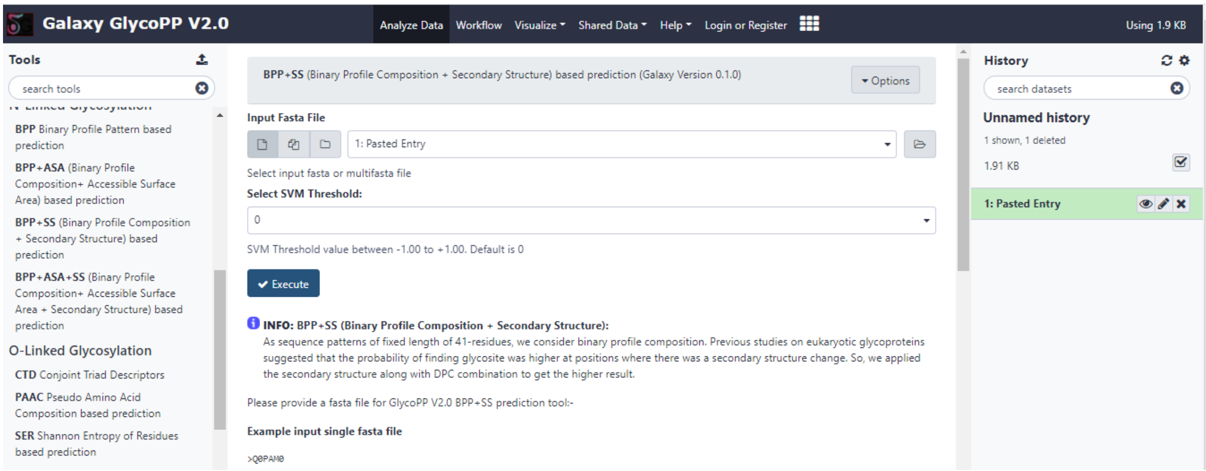
Figure 36: *GlycoPP V2.0 Analsis overview*



```
>Q0PAM0                                        Length = 171
>Q0PAM0                                        Length = 171
Potential N-Linked Glycosylated Sites:
MAIKIFGILIALFTITFTILSLQDPYSLNLQTNALNFKNIEAKNLKAYESNTSIIKAYYKANSWVRYADRDEFNDFITLNLDFNLSANRLEFFNKDM
LFEGNVTYIGSNNVKIISQEVEYQPKDKILHTNTNFKALINGSIINGNTLNYDVKNKILNIQGVNAWLQDK
GlycoPP v2.0 Prediction Method = BPP+ASA+SS    SVM Threshold = 0
------------------------------------------------------------------------------------------------
Position                                       Residue       Score          Prediction
------------------------------------------------------------------------------------------------
29                                             NLQ           -0.61250663    Non-glycosylated
33                                             NAL           -0.53444309    Non-glycosylated
36                                             NFK           0.026839411    Potential Glycosylat
39                                             NIE           -0.29871205    Non-glycosylated
44                                             NLK           -0.30010003    Non-glycosylated
51                                             NTS           0.58642466     Potential Glycosylat
62                                             NSW           -0.39528659    Non-glycosylated
74                                             NDF           -0.20714487    Non-glycosylated
80                                             NLD           -0.65906657    Non-glycosylated
84                                             NLS           0.55723666     Potential Glycosylat
88                                             NRL           -0.28949737    Non-glycosylated
94                                             NKD           -0.23029647    Non-glycosylated
```

Figure 37 GlycoPP V2.0 Analsis output for BPP+ASA+SS (Binary Profile Composition+ Accessible Surface Area + Secondary Structure) forN-linked glycosylation

BPP+SS (Binary Profile Composition+ Secondary Structure) based prediction


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 1.9 KB
Tools
search tools
N-Linked Glycosylation
BPP Binary Profile Pattern based prediction
BPP+ASA (Binary Profile Composition+ Accessible Surface Area) based prediction
BPP+SS (Binary Profile Composition + Secondary Structure) based prediction
BPP+ASA+SS (Binary Profile Composition+ Accessible Surface Area + Secondary Structure) based prediction
O-Linked Glycosylation
CTD Conjoint Triad Descriptors
PAAC Pseudo Amino Acid Composition based prediction
SER Shannon Entropy of Residues based prediction
Options
Execute
History
search datasets
Unnamed history
1 shown, 1 deleted
1.91 KB
1: Pasted Entry

BPP+SS (Binary Profile Composition + Secondary Structure) based prediction (Galaxy Version 0.1.0)

Input Fasta File

1: Pasted Entry

Select input fasta or multifasta file

Select SVM Threshold:

0

SVM Threshold value between -1.00 to +1.00. Default is 0

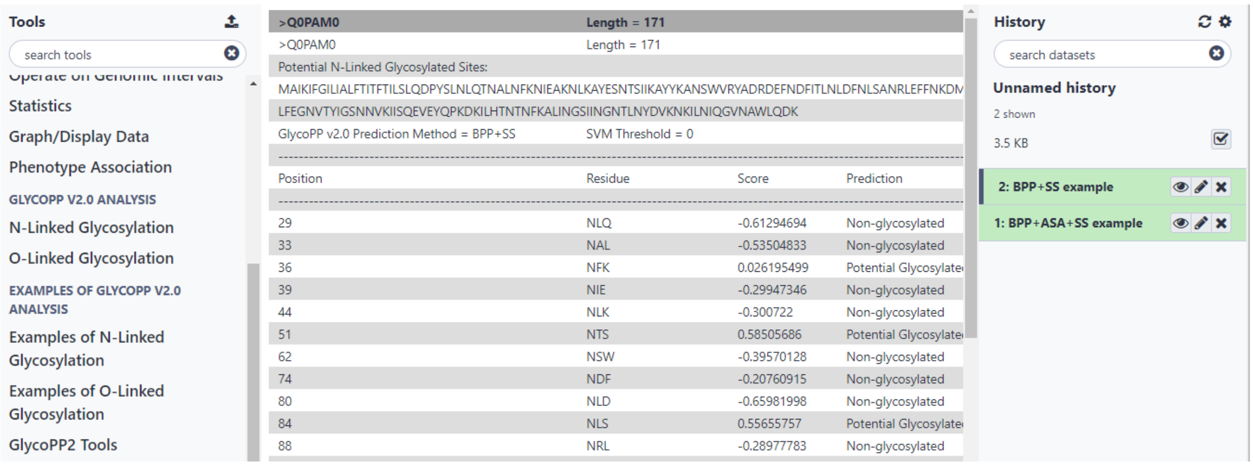
INFO: BPP+SS (Binary Profile Composition + Secondary Structure):
As sequence patterns of fixed length of 41-residues, we consider binary profile composition. Previous studies on eukaryotic glycoproteins suggested that the probability of finding glycosite was higher at positions where there was a secondary structure change. So, we applied the secondary structure along with DPC combination to get the higher result.

Please provide a fasta file for GlycoPP V2.0 BPP+SS prediction tool:-

Example input single fasta file

```
>Q0PAM0
```

Figure 38: *GlycoPP V2.0 Analsis overview*

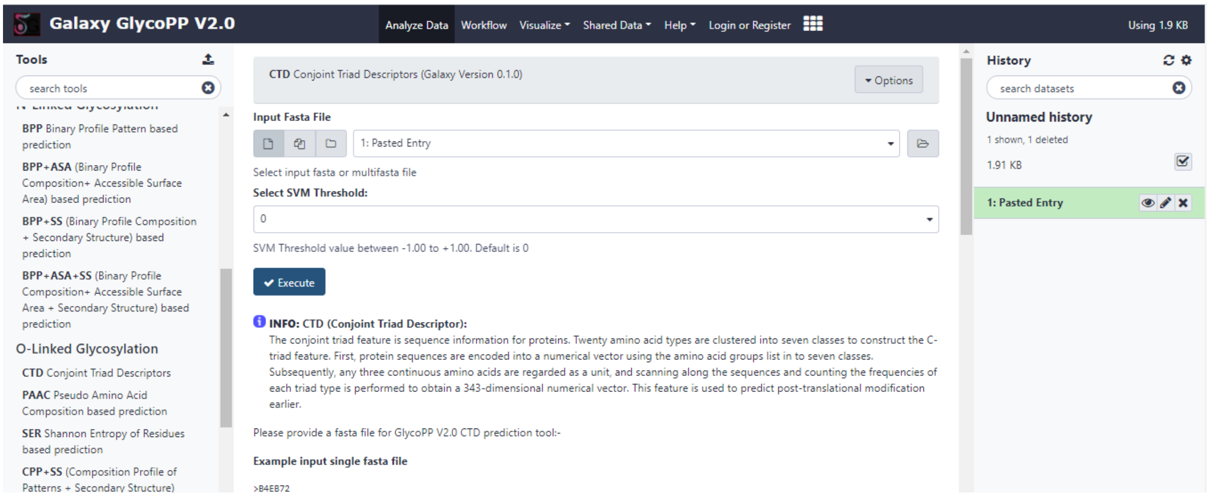


```
>Q0PAM0                                Length = 171
>Q0PAM0                                Length = 171
Potential N-Linked Glycosylated Sites:
MAIKIFGILIALFTITFTILSLQDPYSLNLQTNALNFKNIEAKNLKAYESNTSIIKAYYKANSWVRYADRDEFNDFITLNLDFNLSANRLEFFNKDM
LFEGNVTYIGSNNVKIISQEVEYQPKDKILHTNTNFKALINGSIINGNTLNYDVKNKILNIQGVNAWLQDK
GlycoPP v2.0 Prediction Method = BPP+SS   SVM Threshold = 0
-----------------------------------------------------------------------------------------------------
Position                               Residue      Score          Prediction
-----------------------------------------------------------------------------------------------------
29                                     NLQ          -0.61294694    Non-glycosylated
33                                     NAL          -0.53504833    Non-glycosylated
36                                     NFK          0.026195499    Potential Glycosylate
39                                     NIE          -0.29947346    Non-glycosylated
44                                     NLK          -0.300722      Non-glycosylated
51                                     NTS          0.58505686     Potential Glycosylate
62                                     NSW          -0.39570128    Non-glycosylated
74                                     NDF          -0.20760915    Non-glycosylated
80                                     NLD          -0.65981998    Non-glycosylated
84                                     NLS          0.55655757     Potential Glycosylate
88                                     NRL          -0.28977783    Non-glycosylated
```

Figure 39 *GlycoPP V2.0 Analsis output for BPP+ASA+SS (Binary Profile Composition+ Accessible Surface Area + Secondary Structure) forN-linked glycosylation*

b) O-linked Glycosylation

CTD Conjoint Triad Descriptors


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 1.9 KB
Tools
search tools
N-Linked Glycosylation
BPP Binary Profile Pattern based prediction
BPP+ASA (Binary Profile Composition+ Accessible Surface Area) based prediction
BPP+SS (Binary Profile Composition + Secondary Structure) based prediction
BPP+ASA+SS (Binary Profile Composition+ Accessible Surface Area + Secondary Structure) based prediction
O-Linked Glycosylation
CTD Conjoint Triad Descriptors
PAAC Pseudo Amino Acid Composition based prediction
SER Shannon Entropy of Residues based prediction
CPP+SS (Composition Profile of Patterns + Secondary Structure)
History
search datasets
Unnamed history
1 shown, 1 deleted
1.91 KB
1: Pasted Entry
Options
Execute

CTD Conjoint Triad Descriptors (Galaxy Version 0.1.0)

Input Fasta File

1: Pasted Entry

Select input fasta or multifasta file

Select SVM Threshold:

0

SVM Threshold value between -1.00 to +1.00. Default is 0

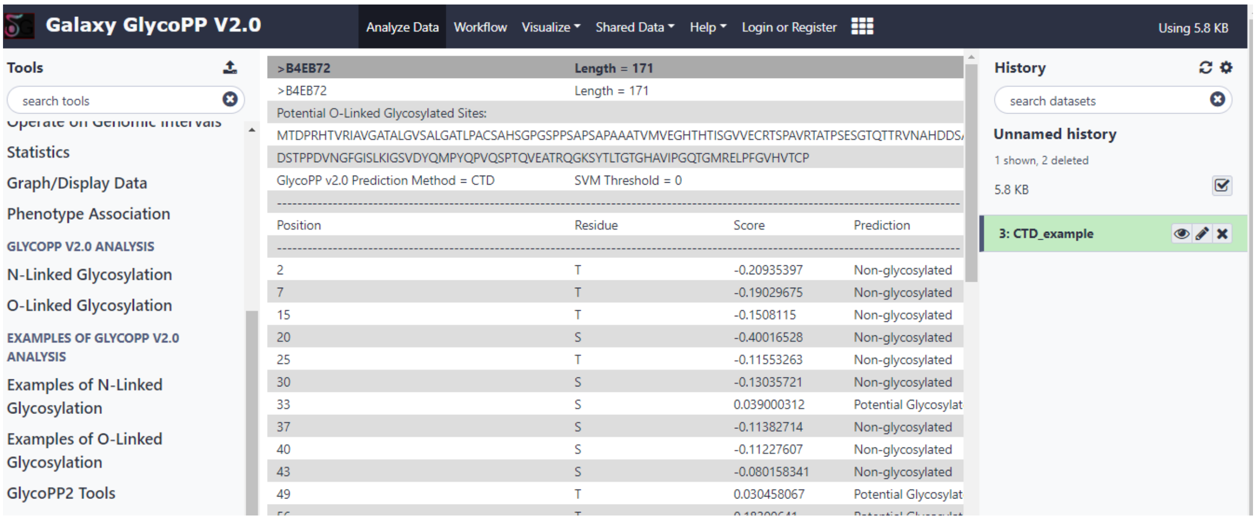
INFO: CTD (Conjoint Triad Descriptor):
The conjoint triad feature is sequence information for proteins. Twenty amino acid types are clustered into seven classes to construct the C-triad feature. First, protein sequences are encoded into a numerical vector using the amino acid groups list in to seven classes. Subsequently, any three continuous amino acids are regarded as a unit, and scanning along the sequences and counting the frequencies of each triad type is performed to obtain a 343-dimensional numerical vector. This feature is used to predict post-translational modification earlier.

Please provide a fasta file for GlycoPP V2.0 CTD prediction tool:-

Example input single fasta file

>B4EB72

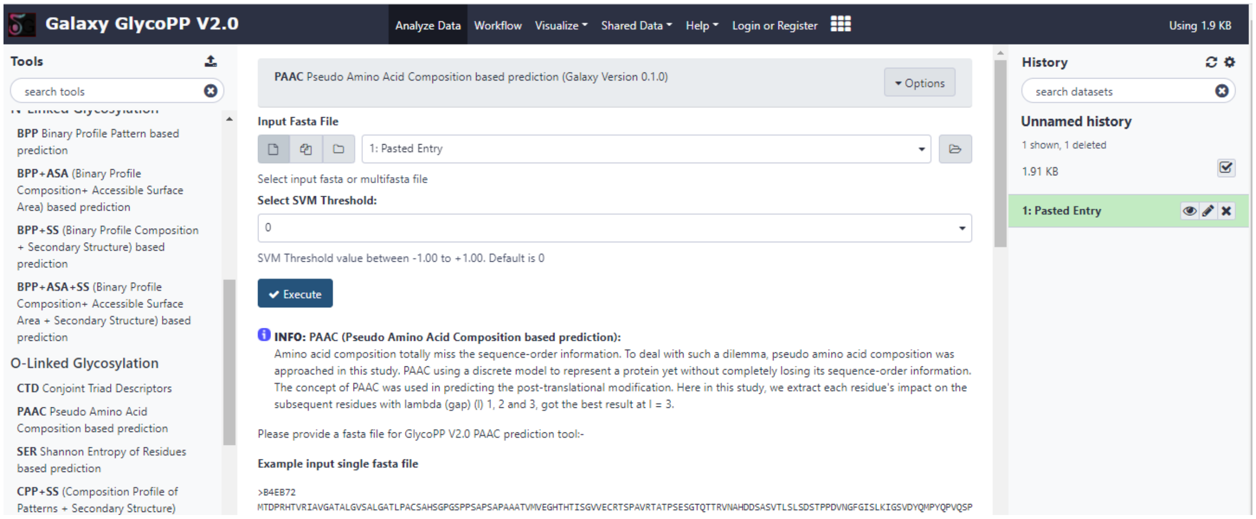
__Figure 40:__ GlycoPP V2.0 Analsis overview



```
>B4EB72                          Length = 171
>B4EB72                          Length = 171
Potential O-Linked Glycosylated Sites:
MTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTISGVVECRTSPAVRTATPSESGTQTTRVNAHDDS/
DSTPPDVNGFGISLKIGSVDYQMPYQPVQSPTQVEATRQGKSYTLTGTGHAVIPGQTGMRELPFGVHVTCP
GlycoPP v2.0 Prediction Method = CTD     SVM Threshold = 0
-----------------------------------------------------------------------------------------------
Position                         Residue          Score           Prediction
-----------------------------------------------------------------------------------------------
2                                T                -0.20935397     Non-glycosylated
7                                T                -0.19029675     Non-glycosylated
15                               T                -0.1508115      Non-glycosylated
20                               S                -0.40016528     Non-glycosylated
25                               T                -0.11553263     Non-glycosylated
30                               S                -0.13035721     Non-glycosylated
33                               S                0.039000312     Potential Glycosylat
37                               S                -0.11382714     Non-glycosylated
40                               S                -0.11227607     Non-glycosylated
43                               S                -0.080158341    Non-glycosylated
49                               T                0.030458067     Potential Glycosylat
```

__Figure 41:__ GlycoPP V2.0 Analsis output for CTD (Conjoint Triad Distributor) forO-linked glycosylation

PAAC Pseudo Amino Acid Composition based prediction


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 1.9 KB
Tools
search tools
N-Linked Glycosylation
BPP Binary Profile Pattern based prediction
BPP+ASA (Binary Profile Composition+ Accessible Surface Area) based prediction
BPP+SS (Binary Profile Composition + Secondary Structure) based prediction
BPP+ASA+SS (Binary Profile Composition+ Accessible Surface Area + Secondary Structure) based prediction
O-Linked Glycosylation
CTD Conjoint Triad Descriptors
PAAC Pseudo Amino Acid Composition based prediction
SER Shannon Entropy of Residues based prediction
CPP+SS (Composition Profile of Patterns + Secondary Structure)
History
search datasets
Unnamed history
1 shown, 1 deleted
1.91 KB
1: Pasted Entry
Options
Execute

PAAC Pseudo Amino Acid Composition based prediction (Galaxy Version 0.1.0)

Input Fasta File

1: Pasted Entry

Select input fasta or multifasta file

Select SVM Threshold:

0

SVM Threshold value between -1.00 to +1.00. Default is 0

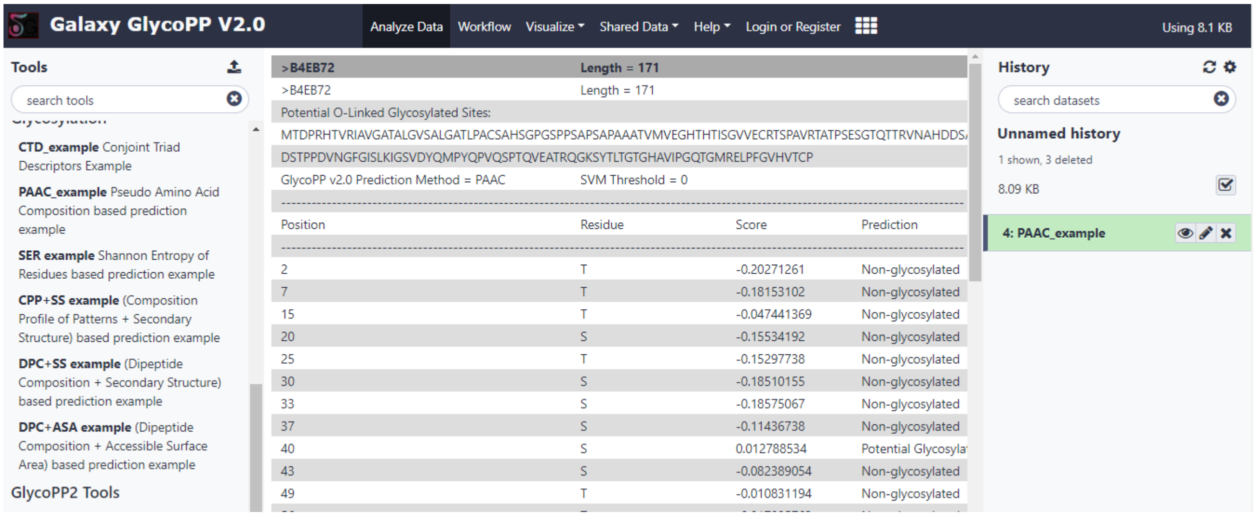
INFO: PAAC (Pseudo Amino Acid Composition based prediction):
Amino acid composition totally miss the sequence-order information. To deal with such a dilemma, pseudo amino acid composition was approached in this study. PAAC using a discrete model to represent a protein yet without completely losing its sequence-order information. The concept of PAAC was used in predicting the post-translational modification. Here in this study, we extract each residue's impact on the subsequent residues with lambda (gap) (l) 1, 2 and 3, got the best result at l = 3.

Please provide a fasta file for GlycoPP V2.0 PAAC prediction tool:-

Example input single fasta file

```
>B4EB72
MTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTISGVVECRTSPAVRTATPSESGTQTTRVNAHDDSASVTLSLSDSTPPDVNGFGISLKIGSVDYQMPYQPVQSP
```

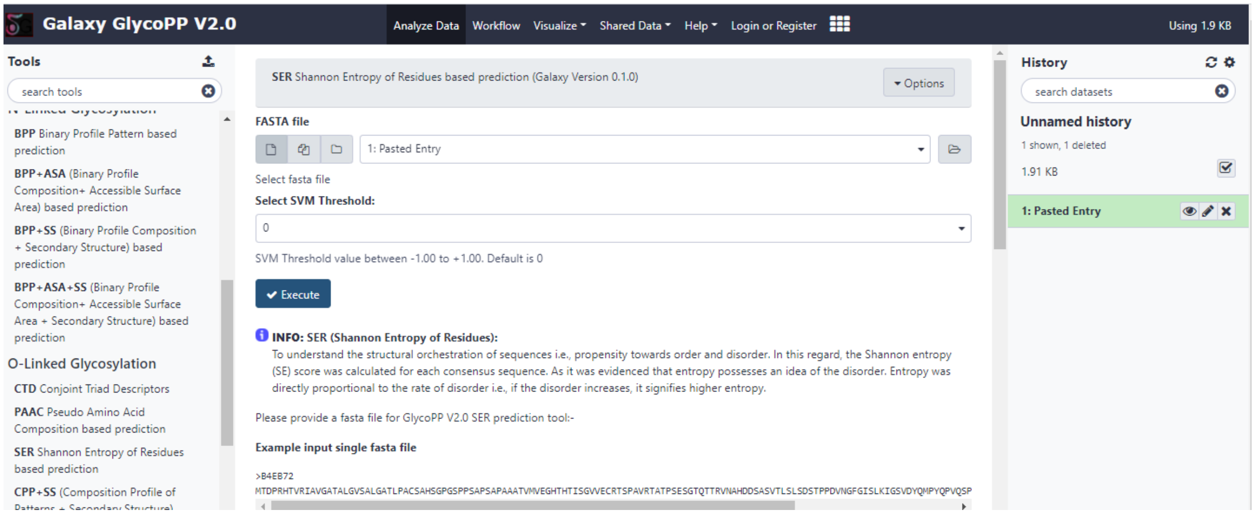
Figure 42: GlycoPP V2.0 Analsis overview


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 8.1 KB
Tools
search tools
Glycosylation
CTD_example Conjoint Triad Descriptors Example
PAAC_example Pseudo Amino Acid Composition based prediction example
SER example Shannon Entropy of Residues based prediction example
CPP+SS example (Composition Profile of Patterns + Secondary Structure) based prediction example
DPC+SS example (Dipeptide Composition + Secondary Structure) based prediction example
DPC+ASA example (Dipeptide Composition + Accessible Surface Area) based prediction example
GlycoPP2 Tools
History
search datasets
Unnamed history
1 shown, 3 deleted
8.09 KB
4: PAAC_example

```
>B4EB72                                   Length = 171
>B4EB72                                   Length = 171
Potential O-Linked Glycosylated Sites:
MTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTISGVVECRTSPAVRTATPSESGTQTTRVNAHDDS/
DSTPPDVNGFGISLKIGSVDYQMPYQPVQSPTQVEATRQGKSYTLTGTGHAVIPGQTGMRELPFGVHVTCP
GlycoPP v2.0 Prediction Method = PAAC     SVM Threshold = 0
--------------------------------------------------------------------------------------------
Position                                  Residue        Score            Prediction
--------------------------------------------------------------------------------------------
2                                         T              -0.20271261      Non-glycosylated
7                                         T              -0.18153102      Non-glycosylated
15                                        T              -0.047441369     Non-glycosylated
20                                        S              -0.15534192      Non-glycosylated
25                                        T              -0.15297738      Non-glycosylated
30                                        S              -0.18510155      Non-glycosylated
33                                        S              -0.18575067      Non-glycosylated
37                                        S              -0.11436738      Non-glycosylated
40                                        S              0.012788534      Potential Glycosyla
43                                        S              -0.082389054     Non-glycosylated
49                                        T              -0.010831194     Non-glycosylated
```

Figure 43: GlycoPP V2.0 Analsis output for PAAC (Pseudo Amino Acid Composition) forO-linked glycosylation

SER Shannon Entropy of Residues based prediction



Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 1.9 KB
Tools
search tools
N-Linked Glycosylation
BPP Binary Profile Pattern based prediction
BPP+ASA (Binary Profile Composition+ Accessible Surface Area) based prediction
BPP+SS (Binary Profile Composition + Secondary Structure) based prediction
BPP+ASA+SS (Binary Profile Composition+ Accessible Surface Area + Secondary Structure) based prediction
O-Linked Glycosylation
CTD Conjoint Triad Descriptors
PAAC Pseudo Amino Acid Composition based prediction
SER Shannon Entropy of Residues based prediction
CPP+SS (Composition Profile of Patterns + Secondary Structure)
Options
History
search datasets
Unnamed history
1 shown, 1 deleted
1.91 KB
1: Pasted Entry

SER Shannon Entropy of Residues based prediction (Galaxy Version 0.1.0)

FASTA file

1: Pasted Entry

Select fasta file

Select SVM Threshold:

0

SVM Threshold value between -1.00 to +1.00. Default is 0

✔ Execute

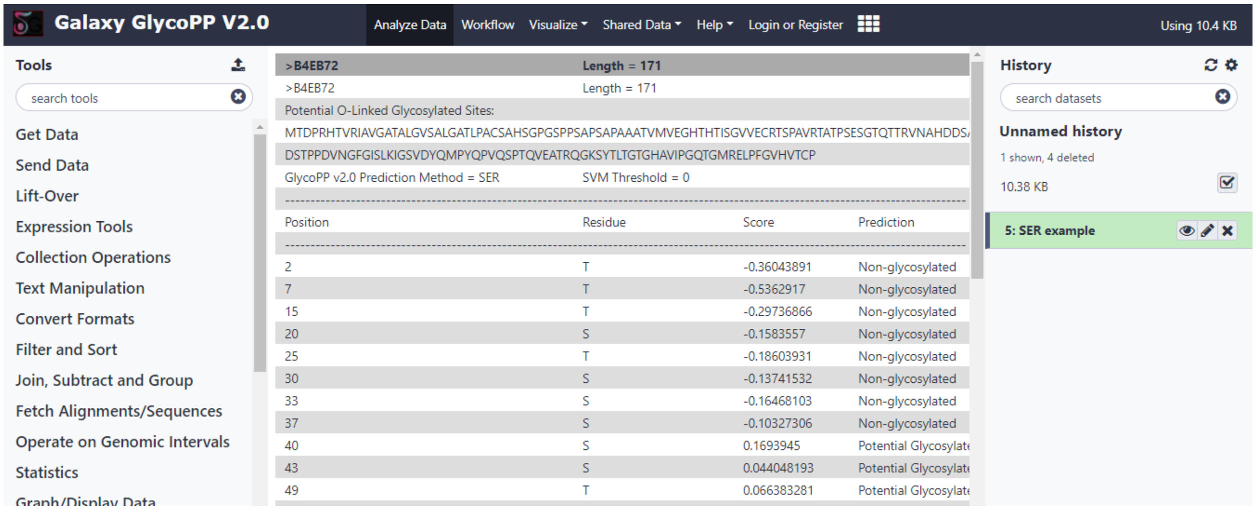
INFO: SER (Shannon Entropy of Residues):
To understand the structural orchestration of sequences i.e., propensity towards order and disorder. In this regard, the Shannon entropy (SE) score was calculated for each consensus sequence. As it was evidenced that entropy possesses an idea of the disorder. Entropy was directly proportional to the rate of disorder i.e., if the disorder increases, it signifies higher entropy.

Please provide a fasta file for GlycoPP V2.0 SER prediction tool:-

Example input single fasta file

```
>B4EB72
MTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTISGVVECRTSPAVRTATPSESGTQTTRVNAHDDSASVTLSLSDSTPPDVNGFGISLKIGSVDYQMPYQPVQSP
```

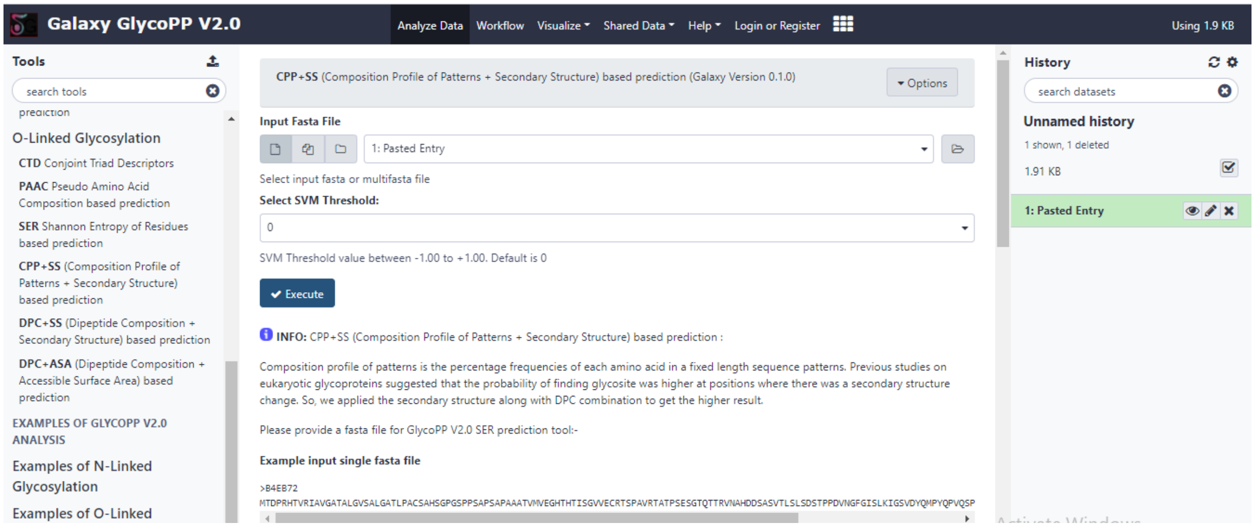
Figure 44: *GlycoPP V2.0 Analsis overview*



```
>B4EB72                                  Length = 171
>B4EB72                                  Length = 171
Potential O-Linked Glycosylated Sites:
MTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTISGVVECRTSPAVRTATPSESGTQTTRVNAHDDS/
DSTPPDVNGFGISLKIGSVDYQMPYQPVQSPTQVEATRQGKSYTLTGTGHAVIPGQTGMRELPFGVHVTCP
GlycoPP v2.0 Prediction Method = SER     SVM Threshold = 0
------------------------------------------------------------------------------------------
Position                                 Residue          Score          Prediction
------------------------------------------------------------------------------------------
2                                        T                -0.36043891    Non-glycosylated
7                                        T                -0.5362917     Non-glycosylated
15                                       T                -0.29736866    Non-glycosylated
20                                       S                -0.1583557     Non-glycosylated
25                                       T                -0.18603931    Non-glycosylated
30                                       S                -0.13741532    Non-glycosylated
33                                       S                -0.16468103    Non-glycosylated
37                                       S                -0.10327306    Non-glycosylated
40                                       S                0.1693945      Potential Glycosylate
43                                       S                0.044048193    Potential Glycosylate
49                                       T                0.066383281    Potential Glycosylate
```

Figure 45: *GlycoPP V2.0 Analsis output for SER (Shannon Entropy of Residues) forO-linked glycosylation*

CPP+SS (Composition Profile of Patterns + Secondary Structure) based prediction


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 1.9 KB
Tools
search tools
prediction
O-Linked Glycosylation
CTD Conjoint Triad Descriptors
PAAC Pseudo Amino Acid Composition based prediction
SER Shannon Entropy of Residues based prediction
CPP+SS (Composition Profile of Patterns + Secondary Structure) based prediction
DPC+SS (Dipeptide Composition + Secondary Structure) based prediction
DPC+ASA (Dipeptide Composition + Accessible Surface Area) based prediction
EXAMPLES OF GLYCOPP V2.0 ANALYSIS
Examples of N-Linked Glycosylation
Examples of O-Linked
Options
History
search datasets
Unnamed history
1 shown, 1 deleted
1.91 KB
1: Pasted Entry
Execute

CPP+SS (Composition Profile of Patterns + Secondary Structure) based prediction (Galaxy Version 0.1.0)

Input Fasta File

1: Pasted Entry

Select input fasta or multifasta file

Select SVM Threshold:

0

SVM Threshold value between -1.00 to +1.00. Default is 0

INFO: CPP+SS (Composition Profile of Patterns + Secondary Structure) based prediction :

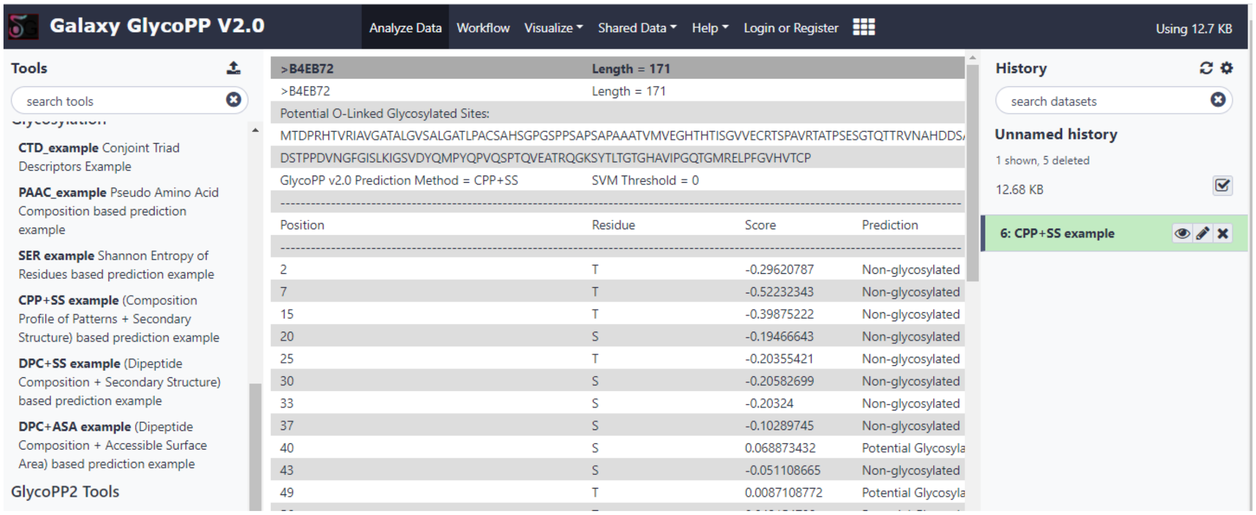
Composition profile of patterns is the percentage frequencies of each amino acid in a fixed length sequence patterns. Previous studies on eukaryotic glycoproteins suggested that the probability of finding glycosite was higher at positions where there was a secondary structure change. So, we applied the secondary structure along with DPC combination to get the higher result.

Please provide a fasta file for GlycoPP V2.0 SER prediction tool:-

Example input single fasta file

```
>B4EB72
MTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTISGVVECRTSPAVRTATPSESGTQTTRVNAHDDSASVTLSLSDSTPPDVNGFGISLKIGSVDYQMPYQPVQSP
```

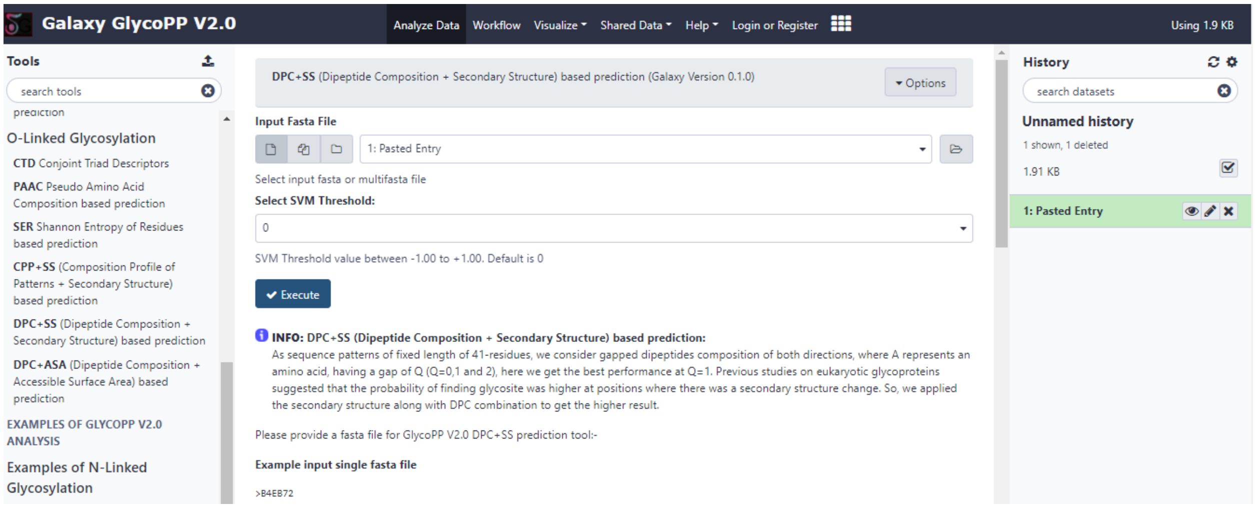
Figure 46: *GlycoPP V2.0 Analsis overview*


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 12.7 KB
Tools
search tools
Glycosylation
CTD_example Conjoint Triad Descriptors Example
PAAC_example Pseudo Amino Acid Composition based prediction example
SER example Shannon Entropy of Residues based prediction example
CPP+SS example (Composition Profile of Patterns + Secondary Structure) based prediction example
DPC+SS example (Dipeptide Composition + Secondary Structure) based prediction example
DPC+ASA example (Dipeptide Composition + Accessible Surface Area) based prediction example
GlycoPP2 Tools
History
search datasets
Unnamed history
1 shown, 5 deleted
12.68 KB
6: CPP+SS example

```
>B4EB72                              Length = 171
>B4EB72                              Length = 171
Potential O-Linked Glycosylated Sites:
MTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTISGVVECRTSPAVRTATPSESGTQTTRVNAHDDS/
DSTPPDVNGFGISLKIGSVDYQMPYQPVQSPTQVEATRQGKSYTLTGTGHAVIPGQTGMRELPFGVHVTCP
GlycoPP v2.0 Prediction Method = CPP+SS    SVM Threshold = 0
------------------------------------------------------------------------------------------
Position                             Residue          Score            Prediction
------------------------------------------------------------------------------------------
2                                    T                -0.29620787      Non-glycosylated
7                                    T                -0.52232343      Non-glycosylated
15                                   T                -0.39875222      Non-glycosylated
20                                   S                -0.19466643      Non-glycosylated
25                                   T                -0.20355421      Non-glycosylated
30                                   S                -0.20582699      Non-glycosylated
33                                   S                -0.20324         Non-glycosylated
37                                   S                -0.10289745      Non-glycosylated
40                                   S                0.068873432      Potential Glycosyla
43                                   S                -0.051108665     Non-glycosylated
49                                   T                0.0087108772     Potential Glycosyla
```

Figure 47: GlycoPP V2.0 Analsis output for CPP+SS (Composition Profile of Patterns + Secondary Structure) forO-linked glycosylation

DPC+SS (Dipeptide Composition + Secondary Structure)based prediction


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 1.9 KB
Tools
search tools
prediction
O-Linked Glycosylation
CTD Conjoint Triad Descriptors
PAAC Pseudo Amino Acid Composition based prediction
SER Shannon Entropy of Residues based prediction
CPP+SS (Composition Profile of Patterns + Secondary Structure) based prediction
DPC+SS (Dipeptide Composition + Secondary Structure) based prediction
DPC+ASA (Dipeptide Composition + Accessible Surface Area) based prediction
EXAMPLES OF GLYCOPP V2.0 ANALYSIS
Examples of N-Linked Glycosylation
Options
History
search datasets
Unnamed history
1 shown, 1 deleted
1.91 KB
1: Pasted Entry

DPC+SS (Dipeptide Composition + Secondary Structure) based prediction (Galaxy Version 0.1.0)

Input Fasta File

1: Pasted Entry

Select input fasta or multifasta file

Select SVM Threshold:

0

SVM Threshold value between -1.00 to +1.00. Default is 0

✓ Execute

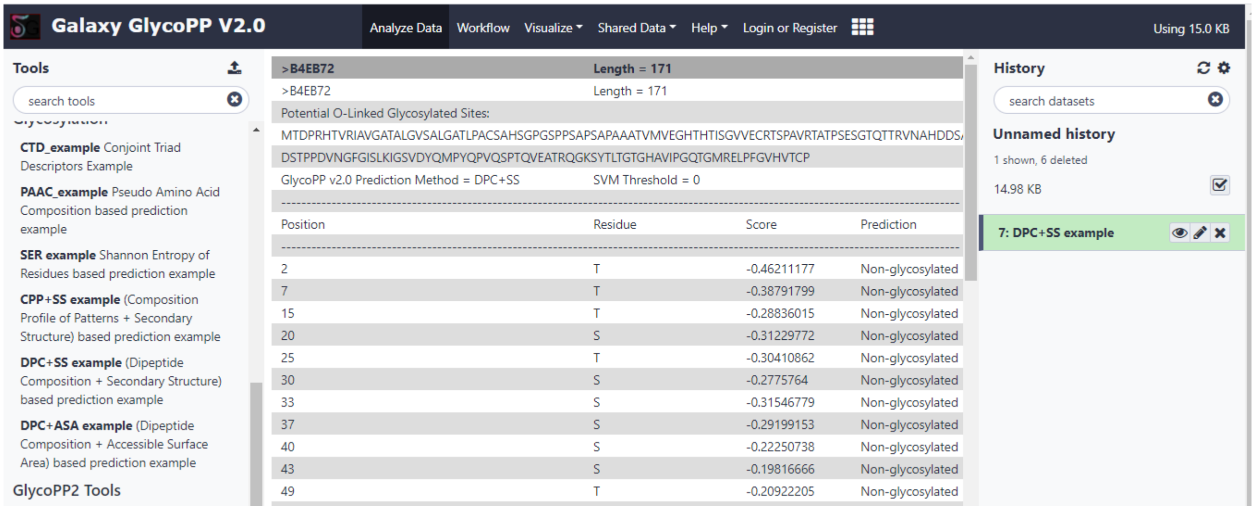
INFO: DPC+SS (Dipeptide Composition + Secondary Structure) based prediction:
As sequence patterns of fixed length of 41-residues, we consider gapped dipeptides composition of both directions, where A represents an amino acid, having a gap of Q (Q=0,1 and 2), here we get the best performance at Q=1. Previous studies on eukaryotic glycoproteins suggested that the probability of finding glycosite was higher at positions where there was a secondary structure change. So, we applied the secondary structure along with DPC combination to get the higher result.

Please provide a fasta file for GlycoPP V2.0 DPC+SS prediction tool:-

Example input single fasta file

>B4EB72

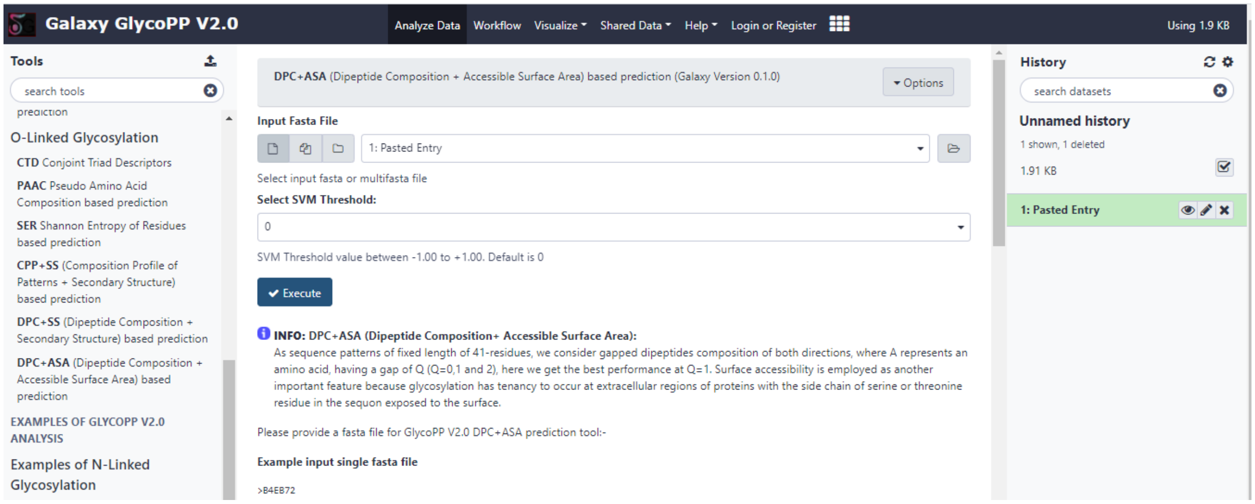
__Figure 48:__ GlycoPP V2.0 Analsis overview


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 15.0 KB
Tools
search tools
Glycosylation
CTD_example Conjoint Triad Descriptors Example
PAAC_example Pseudo Amino Acid Composition based prediction example
SER example Shannon Entropy of Residues based prediction example
CPP+SS example (Composition Profile of Patterns + Secondary Structure) based prediction example
DPC+SS example (Dipeptide Composition + Secondary Structure) based prediction example
DPC+ASA example (Dipeptide Composition + Accessible Surface Area) based prediction example
GlycoPP2 Tools
History
search datasets
Unnamed history
1 shown, 6 deleted
14.98 KB
7: DPC+SS example

```
>B4EB72                                   Length = 171
>B4EB72                                   Length = 171
Potential O-Linked Glycosylated Sites:
MTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTISGVVECRTSPAVRTATPSESGTQTTRVNAHDDS/
DSTPPDVNGFGISLKIGSVDYQMPYQPVQSPTQVEATRQGKSYTLTGTGHAVIPGQTGMRELPFGVHVTCP
GlycoPP v2.0 Prediction Method = DPC+SS   SVM Threshold = 0
-----------------------------------------------------------------------------------------------
Position                                  Residue          Score          Prediction
-----------------------------------------------------------------------------------------------
2                                         T                -0.46211177    Non-glycosylated
7                                         T                -0.38791799    Non-glycosylated
15                                        T                -0.28836015    Non-glycosylated
20                                        S                -0.31229772    Non-glycosylated
25                                        T                -0.30410862    Non-glycosylated
30                                        S                -0.2775764     Non-glycosylated
33                                        S                -0.31546779    Non-glycosylated
37                                        S                -0.29199153    Non-glycosylated
40                                        S                -0.22250738    Non-glycosylated
43                                        S                -0.19816666    Non-glycosylated
49                                        T                -0.20922205    Non-glycosylated
```

__Figure 49:__ GlycoPP V2.0 Analsis output for DPC+SS(Dipeptide Composition + Secondary Structure)forO-linked glycosylation

DPC+ASA (Dipeptide Composition + Accesible surface Area)based prediction


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 1.9 KB
Tools
search tools
prediction
O-Linked Glycosylation
CTD Conjoint Triad Descriptors
PAAC Pseudo Amino Acid Composition based prediction
SER Shannon Entropy of Residues based prediction
CPP+SS (Composition Profile of Patterns + Secondary Structure) based prediction
DPC+SS (Dipeptide Composition + Secondary Structure) based prediction
DPC+ASA (Dipeptide Composition + Accessible Surface Area) based prediction
EXAMPLES OF GLYCOPP V2.0 ANALYSIS
Examples of N-Linked Glycosylation
Options
History
search datasets
Unnamed history
1 shown, 1 deleted
1.91 KB
1: Pasted Entry

DPC+ASA (Dipeptide Composition + Accessible Surface Area) based prediction (Galaxy Version 0.1.0)

Input Fasta File

1: Pasted Entry

Select input fasta or multifasta file

Select SVM Threshold:

0

SVM Threshold value between -1.00 to +1.00. Default is 0

✓ Execute

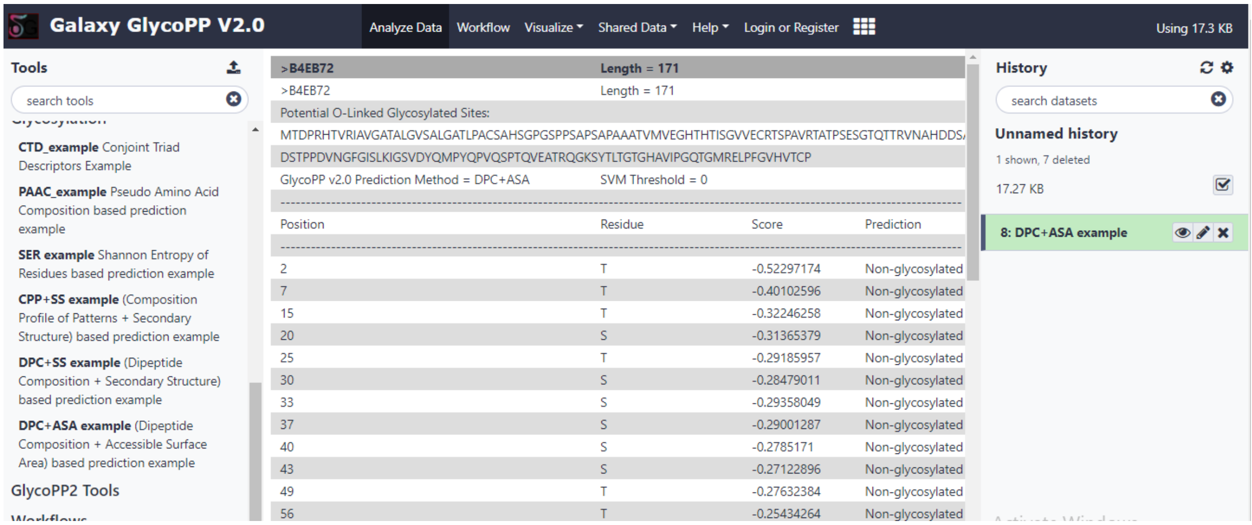
INFO: **DPC+ASA (Dipeptide Composition+ Accessible Surface Area):**
As sequence patterns of fixed length of 41-residues, we consider gapped dipeptides composition of both directions, where A represents an amino acid, having a gap of Q (Q=0,1 and 2), here we get the best performance at Q=1. Surface accessibility is employed as another important feature because glycosylation has tenancy to occur at extracellular regions of proteins with the side chain of serine or threonine residue in the sequon exposed to the surface.

Please provide a fasta file for GlycoPP V2.0 DPC+ASA prediction tool:-

Example input single fasta file

>B4EB72

Figure 50: *GlycoPP V2.0 Analsis overview*


Galaxy GlycoPP V2.0
Analyze Data
Workflow
Visualize
Shared Data
Help
Login or Register
Using 17.3 KB
Tools
search tools
Glycosylation
CTD_example Conjoint Triad Descriptors Example
PAAC_example Pseudo Amino Acid Composition based prediction example
SER example Shannon Entropy of Residues based prediction example
CPP+SS example (Composition Profile of Patterns + Secondary Structure) based prediction example
DPC+SS example (Dipeptide Composition + Secondary Structure) based prediction example
DPC+ASA example (Dipeptide Composition + Accessible Surface Area) based prediction example
GlycoPP2 Tools
Workflows
History
search datasets
Unnamed history
1 shown, 7 deleted
17.27 KB
8: DPC+ASA example

```
>B4EB72                                         Length = 171
>B4EB72                                         Length = 171
Potential O-Linked Glycosylated Sites:
MTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTISGVVECRTSPAVRTATPSESGTQTTRVNAHDDS
DSTPPDVNGFGISLKIGSVDYQMPYQPVQSPTQVEATRQGKSYTLTGTGHAVIPGQTGMRELPFGVHVTCP
GlycoPP v2.0 Prediction Method = DPC+ASA        SVM Threshold = 0
------------------------------------------------------------------------------------------
Position                                        Residue     Score          Prediction
------------------------------------------------------------------------------------------
2                                               T           -0.52297174    Non-glycosylated
7                                               T           -0.40102596    Non-glycosylated
15                                              T           -0.32246258    Non-glycosylated
20                                              S           -0.31365379    Non-glycosylated
25                                              T           -0.29185957    Non-glycosylated
30                                              S           -0.28479011    Non-glycosylated
33                                              S           -0.29358049    Non-glycosylated
37                                              S           -0.29001287    Non-glycosylated
40                                              S           -0.2785171     Non-glycosylated
43                                              S           -0.27122896    Non-glycosylated
49                                              T           -0.27632384    Non-glycosylated
56                                              T           -0.25434264    Non-glycosylated
```

Figure 51: ***GlycoPP V2.0 Analsis output for DPC+ASA (Dipeptide Composition + Accesible surface Area) forO-linked glycosylation***

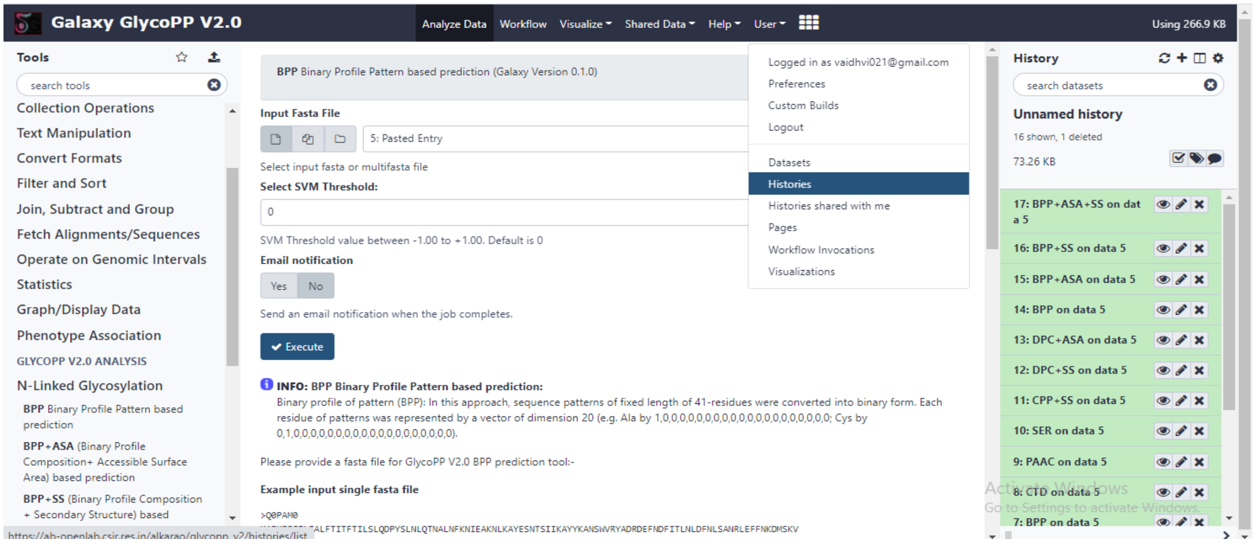
III. Detail Panel

This panel displays the interface of all the tools along with Input Parameters required to run a tool. It also provides help and examples to run a tool. This panel also displays the Output of a tool after its execution when user clicks on the eye 👁 icon show in **History Panel**.

IV. History Panel

When data is uploaded from your computer or analysis is done on existing data using Galaxy, each output from those steps generates a dataset. These datasets (and the output datasets from later analysis on them) are stored by Galaxy in **Histories**.

Users that have registered an account and logged in can have many histories and the history panel allows switching between them and creating new ones.



BPP Binary Profile Pattern based prediction (Galaxy Version 0.1.0)

Input Fasta File

5: Pasted Entry

Select input fasta or multifasta file

Select SVM Threshold:

0

SVM Threshold value between -1.00 to +1.00. Default is 0

Email notification

Yes [✓ No]

Send an email notification when the job completes.

✓ Execute

INFO: BPP Binary Profile Pattern based prediction:

Binary profile of pattern (BPP): In this approach, sequence patterns of fixed length of 41-residues were converted into binary form. Each residue of patterns was represented by a vector of dimension 20 (e.g. Ala by 1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0; Cys by 0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0).

Please provide a fasta file for GlycoPP V2.0 BPP prediction tool:-

Example input single fasta file

```
>Q0PAM0
ALFTITFTILSLQDPYSLNLQTNALNFKNIEAKNLKAYESNTSIIKAYYKANSWVRYADRDEFNDFITLNLDFNLSANRLEFFNKDMSKV
```