	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ
Primary Sequence Analysis of PTMs	Effective Date: 04/26/2021	Revision # 1.0 A. Siclair
Module Hours: 3		
		Checked: M. Guzie

BACKGROUND

The central dogma of biology, emphasizing the transition from DNA to RNA to protein, is fundamentally important when establishing an understanding of protein functionality within cells. However, the genome alone cannot account for the exponentially larger number of proteins that exist compared to the number of genes in the human body. (ThermoFisher Scientific) This observation has led to increased investigation into the importance of numerous posttranslational modifications that allow for the diversification of proteins.

There is an extensive number of posttranslational modifications that can be applied to a given protein, including covalent addition of functional groups and proteolytic cleavage. Many different programs have been developed to analyze primary protein sequences and anticipate where such post translational modifications may occur.

- **PROSITE:** PROSITE is a conglomerate database that contains information about protein families and domains. (G-Preciado et al., 2009) It functions by grouping together proteins with similar domains that indicate similar functions. This site can be used to identify potential post translational modification sites of known proteins or to potentially determine the function of an unknown protein based on similarities to certain families or domains.
- **ProP:** ProP is a server that predicts arginine and lysine propeptide cleavage sites in eukaryotic protein sequences. (Duckert et al., 2004) ProP utilizes artificial neural networks to predict the cleavage sites. (Duckert et al., 2004)
- NetPhos: NetPhos is a server that predicts serine, threonine or tyrosine phosphorylation sites in eukaryotic proteins using artificial neural networks. (Blom et al., 1999).
- **NetOGlyc:** NetOGlyc is a server that predicts GalNAc O-glycosylation sites in mammalian proteins using artificial neural networks. (Steentoft et al, 2013)
- **NetNGlyc:** NetNGlyc is a server that predicts N-glycosylation sites in human proteins using artificial neural networks. (Gupta & Brunak, 2002)

1. PURPOSE

The purpose of this procedure is to understand the basics of post translational modification and to become comfortable utilizing various databases that identify post translational modification sites in the primary protein sequence.

F	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ
Primary Sequence Analysis of PTMs	Effective Date: 04/26/2021	Revision # 1.0 A. Siclair
Module Hours: 3		Checked: M. Guzie

2. SCOPE

This procedure applies to qualified skills center users.

3. **RESPONSIBILITY**

- 3.1. It is the responsibility of the user to understand and perform the procedure described in this document.
- 3.2. It is the responsibility of the user performing the procedure to fully document any deviations from the written procedure.
- 3.3. It is the responsibility of the user to become trained on the procedure.

4. **DEFINITIONS**

- 4.1. Glycosylation The enzymatic, post-translational alteration of a protein involving the covalent addition of a sugar molecule.
- 4.2. Phosphorylation The post-translational addition of a phosphate group, often by a kinase, to specific sites on a protein.
- 4.3. Proteolytic cleavage The hydrolysis of the peptide bonds in a protein that causes the protein to be cut in a specific region. This mechanism is catalyzed by peptidase enzymes.
- 4.4. PROSITE profile A characterization of protein domains over the entire length. A profile is often used to predict the structural properties of a protein and is more sensitive than a PROSITE pattern. (G-Preciado et al., 2009)
- 4.5. PROSITE pattern A short sequence motif confined to a small region with high sequence similarity. A pattern is often used to predict a protein's function. (G-Preciado et al., 2009)
- 4.6. FASTA format A format for representing either nucleotide or protein sequences in which the base pairs or amino acids are represented using the specified single letter codes. (Zhang)

5. MATERIALS/EQUIPMENT

- 5.1. PROSITE application: https://prosite.expasy.org/
- 5.2. ProP application: http://www.cbs.dtu.dk/services/ProP/
- 5.3. NetPhos application: http://www.cbs.dtu.dk/services/NetPhos/
- 5.4. NetOGlyc application: http://www.cbs.dtu.dk/services/NetOGlyc/
- 5.5. NetNGlyc application: http://www.cbs.dtu.dk/services/NetNGlyc/
- 5.6. NCBI Protein database: https://www.ncbi.nlm.nih.gov/protein/

F	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ
Primary Sequence Analysis of PTMs	Effective Date: 04/26/2021	Revision # 1.0 A. Siclair
Module Hours: 3		Checked: M. Guzie

6. PROCEDURE

6.1. PROSITE

- 6.1.1. Go to the PROSITE application website: <u>https://prosite.expasy.org/</u>
- 6.1.2. On the home screen, there are many different options for searching the database. (Figure 2)
 - The Search option is beneficial when there are a few terms, identifiers, etc. that will be used to guide the search.
 - The Browse option is beneficial when there is not as specific of a protein that you are looking for, enabling you to peruse existing entries.
 - The ScanProsite mode is used when searching for information about a protein with a known FASTA sequence, PDB identifier, or UnitProKB accession numbers/identifiers. After pasting the desired input, clicking Scan will bring up a new window with the desired information.
 - For this SOP, the ScanProsite option will be used. The input is the NCBI protein database FASTA sequence for the first hit when searching "human kinase".

Search e.g. PDOC00022, PS50089, SH3, zinc finger Search	Browse by documentation entry by ProRule description by taxonomic scope by number of positive hits	
Quick Scan mode of ScanProsite Quickly find matches of your protein sequences to PROSITE signatures (max. 10 sequences). [?] Examples MNKNDETKIKFKNEDLTDELSLINKISADTTONSGTVIQIJMMANIPEDKILSLLIKLEKNS VPLSDA LIMKLIGNYSQAIEALPPDKYQQNESFARIQVRFAELKAIQEPDDARDYFQMAAAKCKKF AQFELSQBIVKKSKQLLQKAVERGAVPLEHLETALENLINLQKKQLLSEEEKKNLSASTU LTAQES Scan Clear Ø Exclude motifs with a high probability of occurrence from the scan For more scanning options go to ScanProsite	Other tools • PRATT - allows to interactively generate conserved patterns from a series of unaligned proteins. • DyDomains - Image Creator - allows to generate custom domain figures.	

- Clicking the "exclude motifs with a high probability of occurrence" button allows for elimination of common repetitive motifs as hits that can throw off your search. This is most helpful with larger proteins.
- 6.1.3. The ScanPROSITE application will look for hits that are potential matches for the input. The resulting screen first delineates the amino acid sequence of

F	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ PRODUCTON
Primary Sequence Analysis of PTMs	Effective Date: 04/26/2021	Revision # 1.0 A. Siclair
Module Hours: 3		Checked: M. Guzie

the given protein, along with the amount of hits found in X number of sequences. (Figure 3)

prosite	ScanProsite Results Viewer
Ouput format: Graphical view	- this view shows ScanProsite results together with ProRule-based predicted intra-domain features [help].
Hits for all PROSITE (re	elease 2021_02) motifs on sequence USERSEQ1 :
found: 3 hits in 1 sequence	
USERSEQ1 (841 aa)	
MINCVRDIKNIK FKNEDLTDELSLINKT LLINKI LGRYSQAIEALPDRYSQUE FAQFELSGGNWCKSDQL UGAVERG FSGSLGHLQIRNINSCDSRQQTTKAR SPDCDWCTDDSVVPCFMRQTSABE LIITIDSTLINKTESSLAKLEET ARKVITEQKHTTFEQPVFSVSQSQP CULSTPYSQPACFQQQPQLLATPL EKKQIYAIEYMLEEADNQTLDSYM SULKKKKSIDPMERKSYMKIPLEAN SVLVKDSQUGTVMPPEALRDMSSI	ADTTDISGTVIQ INMMINEDNE SLLLKE KINSVPLSDA FARTQMR FÆLKATQEPEDDARDYF QMARAIK KKR FAFVHTS WP LEINE TÆLKINLING (KKRLLS EEKKILLSASTVLTAGES TL VGEINMPOQDAE TGVIRISL BOTINET KOGC PFERAVIVALLN IST SUNDOSKSTE LIKLINSVPRISINSKE FED LVSDEKSS GEVQEPE VPE SINGKAVAKIKSE CITIKUIPAASSIMMQI PFEL 215 TSGIFÖRKSI LGVIRVSTE LIDØFJARQIGSGESSKYFQLN METAVLINL QQHSDKI ITAL VDVE I TDØY I YMMIR GOHI DDLN HETAVLINL QQHSDKI I INT VDVE I TDØY I YMMIR GOHI DDLN HETAVLINL QQHSDKI I LIDØFJAL (DOF GARQIPOTIT BEINKSKSKI SPKSDMSLGCL LIVMTI YGK I FPEQUI TINGES KCC LIKBINGKRIST PEL LIDØFJQQI TINGES

- 6.1.4. Scrolling down the results page will give a more detailed description of the hits found in the database.
 - A legend is given explaining result annotations. (Figure 4)
 - Hit options will be listed by both profile and pattern correlations (see definitions section for more details on the differentiation between these 2 categories).

				*												
disulfide bridge	active site	other 'ran	nges' d	other sites												
lease note that th	e graphical re	presentations of	f domains	displayed h	nereafter are	e for illustra	ative purpo	ses only, a	ind that the	eir colors and	I shapes are	e not inte	inded to indica	te homology or	shared function	on.
or more informati	on about how	hese graphical	represent	ations are c	constructed,	, go to http	s://prosite.e	expasy.org	/mydomaii	ns/.						
its by profile	e 11 hit /hu	1 profile) on		man ¹												
			SHOUN	*****												
nie by preme	o. [1 mit (by	i prome) on	i i seque	encej												
pper case repres	ents match po	sitions, lower ca	ase insert p	positions, a	nd the '-' sy	mbol repre	esents dele	tions relati	ve to the n	natching prof	ile.					
pper case repres	ents match po	sitions, lower ca	ase insert p	positions, a	nd the '-' sy	mbol repre	esents dele	tions relati	ve to the n	natching prof	ile.					
lpper case repres	ents match po	sitions, lower ca	ase insert (300	positions, ai	nd the '-' sy 500	mbol repre	esents dele 700	tions relati 800	ve to the n 900	natching prof	ile.					
Ipper case repres	ents match po	sitions, lower ca	ase insert p	positions, a	nd the '-' sy	mbol repre	osents dele	tions relati	ve to the n	natching prof	ile.					
Ipper case repres	ents match po	sitions, lower ca	ase insert p	positions, a	nd the '-' sy 500	mbol repre	osents dele	tions relati	ve to the n	natching prof	ile.					
Jpper case repres	ents match po	sitions, lower ca	ase insert p 300	400	nd the '-' sy	mbol repre	700	tions relati	900 900 (841 aa)	natching prof	ile.					

- Rolling over the protein kinase domain annotation (Figure 4) in the given example will highlight the region in the amino acid sequence given in Figure 3.
- Clicking on the protein kinase annotation will bring up a new window giving a detailed, generalized description about protein kinase signatures

	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ PRODUCTON
Primary Sequence Analysis of PTMs	Effective Date: 04/26/2021	Revision # 1.0 A. Siclair
Module Hours: 3		Checked: M. Guzie

and profiles. This includes descriptions about the protein family, conserved regions and specific patterns and profiles.

- 6.1.5. Further down the results page there will be information about the resulting domain(s). (Figure 5)
 - There is a description of the amino acid range spanned by the given domain.
 - The score value gives an estimate of how well the given profile matches the profile for the domain in question. (Michael, 2016) A score equal to or above the cut off value indicates a motif occurrence. (Sigrist, 2002)
 - The predicted features section describes features with fulfilled conditions. (de Castro et al., 2006)
 - The absent features section describes features with unfulfilled conditions. (de Castro et al., 2006)

- 775: sco	ore = 43.39	96		
YSILKQ <mark>IGSGGSSKV</mark> KIIRLYDYEITDQYI GIVHSDLKPANFLIV sreNGKSKSKispKS DIPEKDLQDVLKCCL	FQVLN-EKK YMVMECGNI D-GMLKLID DVWSLGCIL KRDPKQRIS	QIYAIKYVN -DLNSWLKK FGIANQMQP YYMTYGKTP IPELLAHPY	LEEADNQTLDSYRNEIAYLNKLQQHSD KKS-IDPWERKSYWKNMLEAVHTIHQH DTTSVVKDSQVGTVWYMPPEAIKdmss FQQIINQISKLHAIIDPN-HEIEFP V	
Predicted featur	es:			
DOMAIN	509	775	Protein kinase	[condition: none]
NP_BIND	515	523	ATP	[condition: <30=K> or <feature:ps00107>]</feature:ps00107>
BINDING	537		ATP	[condition: <30=K>]
Absent feature:				
ACT SITE	631		Proton acceptor	[condition not true: D and <fttag:atp>]</fttag:atp>

6.2. ProP

- 6.2.1. Go to the ProP website: http://www.cbs.dtu.dk/services/ProP/
- 6.2.2. Input the sequence of the protein of interest in the FASTA format, which can be found on the NCBI protein database: https://www.ncbi.nlm.nih.gov/protein/. (Figure 6)
 - Checking the Verbose Output box will show each individual score for all 4 of the artificial neural networks used for computation in addition to the average score. Unchecking the box will just show the average score. (Blom, 2017) (Figure 6)

	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ PRODUCTON
Primary Sequence Analysis of PTMs	Effective Date: 04/26/2021	Revision # 1.0 A. Siclair
Module Hours: 3		Checked: M. Guzie

6.2.3. Click the Submit button to run the query. (Figure 6)

ProP 1.0 Server	
ProP 1.0 server predicts arginine and lysine propeptide cleavag convertase (PC) prediction.	ge sites in eukaryotic protein sequences using
For convenience, this server is integrated with the SignalP serve	r predicting the presence and location of signal
Instructions	c c
SUBMISSION	
Paste a single sequence or several sequences in FASTA format	into the field below:
KLHAIIDPNHEIEFPDIPEKDLQDVLKCCLKRDPKQRISIPELLAH	PYVQIQTHPVNQMAKGTT
MKYVLGQLVGLNSPNSILKAAKTLYEHYSGGESHNSSSSKTFEKKR	GKK //
Submit a file in FASTA format directly from your local disk:	
Choose File No file chosen	Z Include signal population
Verbose output	General PC prediction
- Verbose culput	
Submit Clear fields	
Figure 6: Home page of the P	roP server with the hu
Jane en liens page en lien	

- 6.2.4. A new window will pop up with the results for the search, with each column giving different important information. (Figure 7)
 - Position = shows what position in the propeptide that the R or K residue falls.
 - Residue = shows whether the residue is an R or K.
 - Score = a value between 0.000 to 1.000. The closer the score is to 1.000, the more likely the residue will be followed by a cleavage site. 0.500 is the threshold value; a number above threshold indicates that the R or K is likely followed by a cleavage site. (Blom, 2017)
 - Answer = If the prediction of the site meets or exceeds the threshold score, then the output in this column will be "ProP"
 - Context = shows the R or K in question and the predicted cleavage site, along with the surrounding amino acids.

	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ PRODUCTON
Primary Sequence Analysis of PTMs Module Hours: 3	Effective Date: 04/26/2021	Revision # 1.0 A. Siclair Checked: M. Guzie

****	# Furin-	type	cleavag	e s	ite pre	diction	(Arginine	e/Lysine	res	idues) *****
	-		0.050	,						_	
equence	3	K	0.059	5	0.0/8	0.054	0.029	0.075	2	•	MINK VR
equence	5	R	0.16/	5	0.131	0.237	0.181	0.120	2		PINKVK [D]
equence	8	K	0.135	(0.145	0.132	0.151	0.112	2	1	NKAKDIK INK
equence	10	K	0.089	5	0.090	0.097	0.097	0.0/1	21	•	VRDIKNKIFK
equence	12	K	0.060	Ş	0.085	0.052	0.038	0.065	2	•	DIKNKEKINE
equence	24	K	0.053	9	0.071	0.051	0.033	0.059)	•	DELSLNKIIS
equence	55	K	0.065	(0.086	0.072	0.043	0.058)		WLSLLLKILE
equence	58	K	0.054	(0.079	0.044	0.031	0.061)	•	LLLKLEKINS
equence	70	K	0.073	(0.098	0.053	0.077	0.065)		SDALLNKILI
equence	74	R	0.112	(0.127	0.085	0.117	0.117)		LNKLIGR YS
equence	86	K	0.058	(0.086	0.048	0.032	0.064)		EALPPDK YG
equence	95	R	0.075	(0.096	0.062	0.060	0.082)		QNESFAR IQ
equence	99	R	0.155	(0.104	0.147	0.243	0.128)		FARIQVR FA
equence	104	K	0.066	(0.085	0.050	0.059	0.070)		VRFAELK AI
equence	113	R	0.111	(0.099	0.099	0.114	0.133)		QEPDDAR DY
equence	120	R	0.094	(0.103	0.082	0.080	0.110)		DYFQMAR AN
equence	124	K	0.064	(0.080	0.060	0.045	0.070)		MARANCK KF
equence	125	K	0.180	(0.121	0.124	0.250	0.224)		ARANCKK FA
equence	144	K	0.061	(0.079	0.054	0.037	0.075)		LSQGNVK KS

- 6.2.5. A graph summarizing the data might also be available, depending on the protein. (Figure 8)
 - It is important to note that the horizontal grey line indicates the threshold score. Any predictions that rise past this bar are likely candidates for peptide cleavage.



6.3. NetPhos

- 6.3.1. Go to the NetPhos website: http://www.cbs.dtu.dk/services/NetPhos/
- 6.3.2. Input the sequence of the protein of interest in the FASTA format. (Figure
- 6.3.3. The desired residues can be filtered to serine, threonine, tyrosine or a combination of all three. (Figure 9)

F	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ PRODUCTON
Primary Sequence Analysis of PTMs	Effective Date: 04/26/2021	Revision # 1.0 A. Siclair
Module Hours: 3		Checked: M. Guzie

6.3.4. The search can be filtered based on score. Changing the value to 0.5 will ensure that all hits shown will be positive predictions. (Figure 9) (Blom, 2017)
6.3.5. Click submit when the settings are as desired. (Figure 9)

	Instructions		
SUBMISSION			
Paste a single seque	nce or several sequences in <u>FASTA</u> format	into the field below:	
KLHAIIDPNHEIEFP EE MKYVLGQLVGLNSPN	DIPEKDLQDVLKCCLKRDPKQRISIPELLAH SILKAAKTLYEHYSGGESHNSSSSKTFEKKR	PYVQIQTHPVNQMAKGTT *	
Submit a file in FAST	A format directly from your local disk:		
Choose File No fil	e chosen		
Residues to predict	Serine Othreonine Otyrosine 💿 a	II three	
For each residue di	splay only the best prediction		
Display only the sco	ores higher than 0		
Output format 💿 d	lassical O GFF		
Generate graphics			
Submit Clear fiel	ds		
	Jama page of the Na	tDhoo convor wit	h the human kinese EASTA inn

- 6.3.6. A new window will pop up with the results for the search, with each column giving different important information. (Blom, 2017) (Figure 10)
 - # = shows the position of the potential phosphorylation residue
 - x = shows the amino acid one letter code for the potential phosphorylation residue
 - Context = shows the amino acid one letter codes for the sequence surrounding the potential phosphorylation residue, with the residue centered.
 - Score = a value between 0.000 to 1.000. The closer the score is to 1.000, the more likely the residue is to be a true phosphorylation site. 0.500 is the threshold value, and indicates that the prediction is a strong potential candidate.
 - Kinase = prediction of the specific kinase that would phosphorylate the potential phosphorylation residue
 - Answer = a value of YES will be given if the prediction has a score above threshold and meets the criteria of a strong prediction.



#					
<pre># netphos-3.1</pre>	b prediction resu	ults			
#					
# Sequence	# >	Context	Score	Kinase	Answer
#					
# Sequence	17 1	NEDLTDELS	0.571	CKII	YES
# Sequence	17 1	NEDLTDELS	0.441	р38МАРК	•
# Sequence	17 1	NEDLTDELS	0.438	GSK3	•
# Sequence	17 1	NEDLTDELS	0.394	CaM-II	
# Sequence	17 1	NEDLTDELS	0.366	CKI	
# Sequence	17 1	NEDLTDELS	0.355	DNAPK	
# Sequence	17 1	NEDLTDELS	0.349	cdc2	
# Sequence	17 1	NEDLTDELS	0.311	ATM	
# Sequence	17 1	NEDLTDELS	0.264	PKG	
# Sequence	17 1	NEDLTDELS	0.189	RSK	
# Sequence	17 1	NEDLTDELS	0.148	cdk5	
# Sequence	17 1	NEDLTDELS	0.131	PKA	
# Sequence	17 1	NEDLTDELS	0.092	PKB	
# Sequence	17 1	NEDLTDELS	0.089	PKC	
# Sequence	17 1	NEDLTDELS	0.065	unsp	
#					
# Sequence	21 9	5 TDELSLNKI	0.496	cdc2	
# Sequence	21 9	5 TDELSLNKI	0.458	DNAPK	
# Sequence	21 9	5 TDELSLNKI	0.448	CKI	

- 6.3.7. Scrolling to the bottom of the output page will show 2 new viewing formats. (Figure 11)
 - First, there will be an amino acid sequence, with residue position labels along the right side. All amino acids except for the potential phosphorylation sites will be ommitted.
 - Second, there will be a summary graph that allows for visualization of the different potential phosphorylation sites. Each type of phosphorylation site (serine, threonine or tyrosine) has its own color. It is important to note that the pink line indicates the threshold score.



F	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ PRODUCTON
Primary Sequence Analysis of PTMs Module Hours: 3	Effective Date: 04/26/2021	Revision # 1.0 A. Siclair Checked: M. Guzie

6.4. NetOGlyc

- 6.4.1. Go to the NetOGlyc website: http://www.cbs.dtu.dk/services/NetOGlyc/
- 6.4.2. Input the sequence of the protein of interest in the FASTA format and then click submit to begin the query.
- 6.4.3. A new window will pop up with the results for the search, with each column giving different important information. (Figure 12)
 - Strand/frame columns = give the position of the glycosylated residue.
 - Comment = gives the score for the residue, a value between 0.000 to 1.000. The closer the score is to 1.000, the more likely the residue is to be a true glycosulation site. 0.500 is the threshold value, and indicates that the prediction is a strong potential candidate.
 - Comment cont. = a value of POSITIVE will be given if the prediction has a score above threshold and meets the criteria of a good potential prediction.

#date 21-4-26	in necocity c 4.0.0.15						
Type Protein							
segname	source feature start	end score	strand	frame	comment		
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	5	5	0.667043	14	. #POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	6	6	0.394244		
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	9	9	0.434316		
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	19	19	0.325824		
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	25	25	0.277496		
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	33	33	0.273051		
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	37	37	0.232814		
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	38	38	0.480655		
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	41	41	0.262801		
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	46	46	0.5 .		#POSITIVE
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	47	47	0.266134		
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD	54	54	0.0224482		

6.5. NetNGlyc

- 6.5.1. Go to the NetNGlyc website: http://www.cbs.dtu.dk/services/NetNGlyc/
- 6.5.2. Input the sequence of the protein of interest in the FASTA format.
 - By clicking the Show Additional Thresholds box, there is opportunity to create higher or lower confidence levels of the resulting sites. (Figure 13) (Gupta, 2017)
- 6.5.3. Click submit to begin the query.

F	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ PRODUCTON
Primary Sequence Analysis of PTMs	Effective Date: 04/26/2021	Revision # 1.0 A. Siclair
Module Hours: 3		Checked: M. Guzie

NetNGlyc 1.0 S	Server		
The NetNglyc server predi	cts N-Glycosylation sites in human proteins using artificial neural network		
	Instructions		
SUBMISSION			
Paste a single sequence o	r several sequences in FASTA format into the field below:		
KLHAIIDPNHEIEFPDIPE EE MKYVLGQLVGLNSPNSILK	KDLQDVLKCCLKRDPKQRISIPELLAHPYVQIQTHPVNQMAKGTT		
Submit a file in FASTA for Choose File No file cho	nat directly from your local disk: isen		
Alternatively, type in Swiss	-Prot ID/AC (e.g. CBG_HUMAN)		
Generate graphics	Show additional thresholds (0.32, 0.75, 0.90) in the graph(s)		
By default, predictions are	done only on the Asn-Xaa-Ser/Thr sequons (incl. Asn-Pro-Ser/Thr)		
Predict on all Asn re	sidues - use this only if you know what you are doing!		
Submit Clear fields			
Figure 13: H	lome page of the NetNGlyc serve	with the human	kinase FASTA inn

- 6.5.4. A new window will pop up with the results for the search, with each column giving different important informatio n. (Figure 14)
 - First, the amino acid sequence is given. Blue highlighted regions indicate Asn-Xaa-Ser/Thr sequence. Red highlighted regions indicate Asn residues that are predicted to be N-glycosylated. (Gupta, 2017)

222					
80					
160					
240					
320					
400	(Threshold	=0.5)			
480					
560	SegName	Position	Potential	Jury	N-Glvc
640	0000000			agreement	result
720					
800	Sequence	90 NESE	0.4942	(5/9)	
	Sequence	186 NLSA	0.5949	(6/9)	•
80	Sequence	210 NNSC	0.5364	(4/9)	+
160	Sequence	247 NKTK	0.6433	(7/9)	•
240	Sequence	303 NDSC	0.5442	(6/9)	+
320	Sequence	342 NKTE	0,4897	(4/9)	
400	Sequence	546 NOTL	0,6007	(7/9)	•
480	Sequence	827 NSSS	0.4676	(5/9)	2
560					
640					
720					
800					
880					
	89 160 240 320 480 560 640 720 800 160 240 800 160 240 400 480 560 640 720 800 880	80 160 240 320 400 (Threshold 560 5800	80 160 240 320 400 480 560 640 560 560 560 560 560 560 560 560 560 560 560 560 560 <	80 160 240 320 400 480 560 564 560 <	80 160 240 320 480 (Threshold=0.5) 480 480 560 560 560 560 560 560 560 56

- In the table format, like before, values like position and score are given.
- N-Glyc result column annotations for glycosylation sites: (Gupta, 2017)
 - 1. = negative site
 - 2. + = threshold score above 0.5
 - 3. ++ = threshold score above 0.5 and jury agreement 9/9, or threshold score above 0.75
 - 4. +++ = threshold score above 0.75 and jury agreement 9/9
 - 5. ++++ = threshold score above 0.9 and jury agreement 9/9

F	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ PRODUCTON
Primary Sequence Analysis of PTMs	Effective Date: 04/26/2021	Revision # 1.0 A. Siclair
Module Hours: 3		Checked: M. Guzie

- 6.5.5. There will be a summary graph that allows for visualization of the different potential glycosylation sites. It is important to note that the red line indicates the usual 0.5 threshold score. (Figure 15)
 - Additional dotted grey lines indicate the other potential threshold scores, if the show additional thresholds option was chosen in step 6.5.2.



7. TROUBLE SHOOTING

8. REFERENCES

- Blom, N. (2017, Jan 2). *Output Format.* DTU Bioinformatics. <u>http://www.cbs.dtu.dk/services/ProP/output.php</u>
- Blom, N. (2017, Jan 2). Usage instructions. DTU Bioinformatics. http://www.cbs.dtu.dk/services/ProP/instructions.php
- Blom, N. (2017, Jan 5). *NetPhos 3.1 usage instructions*. DTU Bioinformatics. http://www.cbs.dtu.dk/services/NetPhos/instructions.php
- Blom, N. (2017, Jan 5). *Output Format*. DTU Bioinformatics. http://www.cbs.dtu.dk/services/NetPhos-2.0/output.php
- Blom N., Gammeltoft S. & Brunak S. (1999) Sequence and structure based prediction of eukaryotic protein phosphorylation sites. *Journal of Molecular Biology, 294*(5), 1351-1362. <u>http://www.cbs.dtu.dk/services/NetPhos/</u>
- de Castro E., Sigrist C.J.A., Gattiker A., Bulliard V., Petra S. Langendijk-Genevaux P.S., Gasteiger E. Bairoch A. & Hulo N. (2006). ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Research, 34*, W362-W365.

	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ PRODUCTON
Primary Sequence Analysis of PTMs Module Hours: 3	Effective Date: 04/26/2021	Revision # 1.0 A. Siclair Checked: M. Guzie

https://doi.org/10.1093/nar/gkl124

- Duckert P., Brunak S. & Blom, N. (2004) Prediction of proprotein convertase cleavage sites. *Protein Engineering, Design and Selection, 17,* 107-112. http://www.cbs.dtu.dk/services/ProP/
- G-Preciado, A., Peimbert, M. & Marino, E. (2009). Genome Sequence Databases: Types of Data and Bioinformatic Tools. *Encyclopedia of Microbiology (Third Edition),* 211-236. https://doi.org/10.1016/B978-012373944-5.00027-4
- Gupta, R. (2017, Jan 5). *Instructions.* DTU Bioinformatics. http://www.cbs.dtu.dk/services/NetNGlyc/instructions.php
- Gupta, R. (2017, Jan 5) *Output Format.* DTU Bioinformatics. http://www.cbs.dtu.dk/services/NetNGlyc/output.php
- Gupta, R. & Brunak, S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput.* 310-322. https://pubmed.ncbi.nlm.nih.gov/11928486/
- Hulo N., Bairoch A., Bulliard V., Cerutti L., De Castro E., Langendijk-Genevaux P.S., Pagni M. & Sigrist C.J.A. (2006). The PROSITE database. *Nucleic Acid Research, 34*, D227-D230. <u>https://pubmed.ncbi.nlm.nih.gov/16381852/</u>
- Joshi, H. (2017, Jan 3). *Output Format.* DTU Bioinformatics. http://www.cbs.dtu.dk/services/NetOGlyc-4.0/output.php
- ThermoFisher Scientific. (n.d.) Overview of Post-Translational Modifications (PTMs). ThermoFisher Scientific: Protein Biology Resource Library. <u>https://www.thermofisher.com/us/en/home/life-science/protein-biology/</u> protein-biology-learning-center/protein-biology-resource-library/pierce-proteinmethods/overview-post-translational-modification.html
- Sigrist C.J.A., Cerutti L., Hulo N., Gattiker A., Falquet L., Pagni M., Bairoch A. & Bucher P. (2002). PROSITE: a documented database using patterns and profiles as motif descriptors. *Brief Bioinform, 3*(3), 265-74. doi: 10.1093/bib/3.3.265.
- Steentoft C., Vakhrushev S.Y., Joshi H.J., Kong Y., Vester-Christensen M.B., Schjoldager K.T., Lavrsen K., Dabelsteen S., Pedersen N.B., Marcos-Silva L., Gupta R., Bennett E.P., Mandel U., Brunak S., Wandall H.H., Levery S.B. & Clausen H. (2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J*, 32(10),1478-88. 10.1038/emboj.2013.79.
- Steward, K. (2019, Sep 26). Amino Acids the Building Blocks of Proteins. Applied Sciences from Technology Networks. <u>https://www.technologynetworks.com</u> /applied-sciences/articles/essential-amino-acids-chart-abbreviations-andstructure-324357

F	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ PRODUCTON
Primary Sequence Analysis of PTMs Module Hours: 3	Effective Date: 04/26/2021	Revision # 1.0 A. Siclair Checked: M. Guzie

Zhang, Y. (n.d.) *What is FASTA format?* Zhang Lab, University of Michigan. <u>https://zhanglab.dcmb.med.umich.edu/FASTA/</u>

9. MODULE MASTERY TASK

This task will test your ability to analyze the post-translational modifications of a protein of interest.

Finding a Protein of Interest (do not use the human kinase of the SOP)

- 1. What is your protein of interest and why did you choose this protein?
- 2. In one sentence, describe your protein of interest and its function.
- 3. Go to the NCBI database and obtain the FASTA sequence of your protein of interest. Paste that FASTA sequence here.

<u>PROSITE</u>

- 1. How many hits were found in how many sequences?
- 2. How many profile vs. pattern hits does your protein have?
- 3. Is there an annotated domain on your protein?
- 4. Give a brief description of the domain.
- 5. What range of amino acids does the domain from question 3 cover?
- 6. What is the score value for your domain?
- 7. Were there any absent features in the hit for this domain?

ProP

Pick a resulting hit that to answer the following questions.

- 1. What is the position of the residue/cleavage site?
- 2. Does the cleavage site follow an R or a K residue?
- 3. Document the context of the cleavage site.
- 4. What is the score of this hit? What does this indicate?
- 5. What would you conclude about the predictive strength of this hit?
- 6. Submit the summary graph of your protein's hits.

<u>NetPhos</u>

Pick a resulting hit that to answer the following questions.

- 1. What is the position of the potential phosphorylation site?
- 2. Which residue occurs at this phosphorylation site?
- 3. Document the context of the phosphorylation site.

	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ
Primary Sequence Analysis of PTMs	Effective Date: 04/26/2021	Revision # 1.0 A. Siclair
Module Hours: 3		Checked: M. Guzie

- 4. What is the score of this hit? What does this indicate?
- 5. What would you conclude about the predictive strength of this hit?
- 6. Submit the summary graph of your protein's hits.

<u>NetOGlyc</u>

Pick a resulting hit that to answer the following questions.

- 1. What is the position of the potential glycosylation site?
- 2. What is the score of this hit? What does this indicate?
- 3. What would you conclude about the predictive strength of this hit?

<u>NetNGlyc</u>

Pick a resulting hit that to answer the following questions.

- 1. How many Asn-Xaa-Ser/Thr sequons are there in your protein? How many Asn residues are predicted to be N-glycosylated?
- 2. How many "+" hits were there for your protein? How many negative hits?
- 3. Submit the summary graph of your protein's hits.