

List of POI characteristics and predictions

Target name/Identifier:	fl69ORF_0459919
No of AA:	160
MW:	18.1 kDa
Theor. pI:	9.24
Theor. Ext. coeff:	1.37
Uniprot BLAST: (see page 2)	M5A8F1 (Suppressyn, <i>H. sapiens</i> , <u>secreted</u> )
Signal peptide: (see page 3)	putative (AA 1 – 39)
Transmembrane segments: (see page 4)	putative or signal peptide (AA 1 – 39)
Predicted ECD: (see page 4)	putative (AA 40 -160, see below in red)
Unique protease sies: (see page 5)	- Pos. 126 (Asp-N endopeptidase)
Furin sites: (see page 6)	None
Glycosylation sites: (see page 7 + 8)	None (O-glycosylation) None (N-glycosylation)
Antigenicity: (see page 9)	8 antigenic determinants








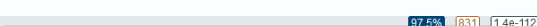



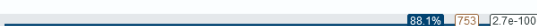
POI sequence

10	20	30	40	50	60
MACIYPTTFY TSLPTKSLNM GISLTILIL SVAVLLSTA <u>PPSCRECYOS LHYRGEMOQY</u>					
70	80	90	100	110	120
<u>FTYHTHIERS CYGNLIEECV ESGKSYKVK NLGVCGRNG AICPRGKOWL CETKIGOWGV</u>					
130	140	150	160		
<u>NTOVLEDIKR EQIIAKAKAS KPTTPPENRP RHFHSFIQKL</u>					

Underlined letters:	“Empirical ECD” as communicated by client (AA 41 – 160)
Red letters:	ECD predicted by trenzyme (AA 40 – 160)

Targetname/Identifier: fl690RF\_0459919

Uniprot blast  
<https://www.uniprot.org/blast/>

Entry	Entry Name	Protein Names	Gene Names	Organism	Length	
<input type="checkbox"/> M5A8F1	 SUPYN_HUMAN	Suppressyn[...]	ERVH48-1, C21orf105, HERV-Fb1, NDUFV3-AS1	Homo sapiens (Human)	160 AA	 100% [864] [1.4e-117]
<input type="checkbox"/> A0A2I2YL90	 A0A2I2YL90_GORGO	Endogenous retrovirus group 48, member 1	ERVH48-1	Gorilla gorilla gorilla (Western lowland gorilla)	160 AA	 99.4% [859] [8.2e-117]
<input type="checkbox"/> A0A2R8Z8Y2	 A0A2R8Z8Y2_PANPA	Endogenous retrovirus group 48 member 1, envelope		Pan paniscus (Pygmy chimpanzee) (Bonobo)	158 AA	 97.5% [831] [1.4e-112]
<input type="checkbox"/> A0A2I3T5X7	 A0A2I3T5X7_PANTR	Endogenous retrovirus group 48 member 1, envelope	ERVH48-1	Pan troglodytes (Chimpanzee)	158 AA	 97.5% [831] [1.4e-112]
<input type="checkbox"/> A0A2I3GT90	 A0A2I3GT90_NOMLE	Endogenous retrovirus group 48 member 1, envelope	ERVH48-1	Nomascus leucogenys (Northern white-cheeked gibbon) (Hylobates leucogenys)	160 AA	 92.5% [787] [7.8e-105]
<input type="checkbox"/> A0A1D5QA25	 A0A1D5QA25_MACMU	Uncharacterized protein	ERVH48-1	Macaca mulatta (Rhesus macaque)	184 AA	 88.1% [753] [2.7e-100]



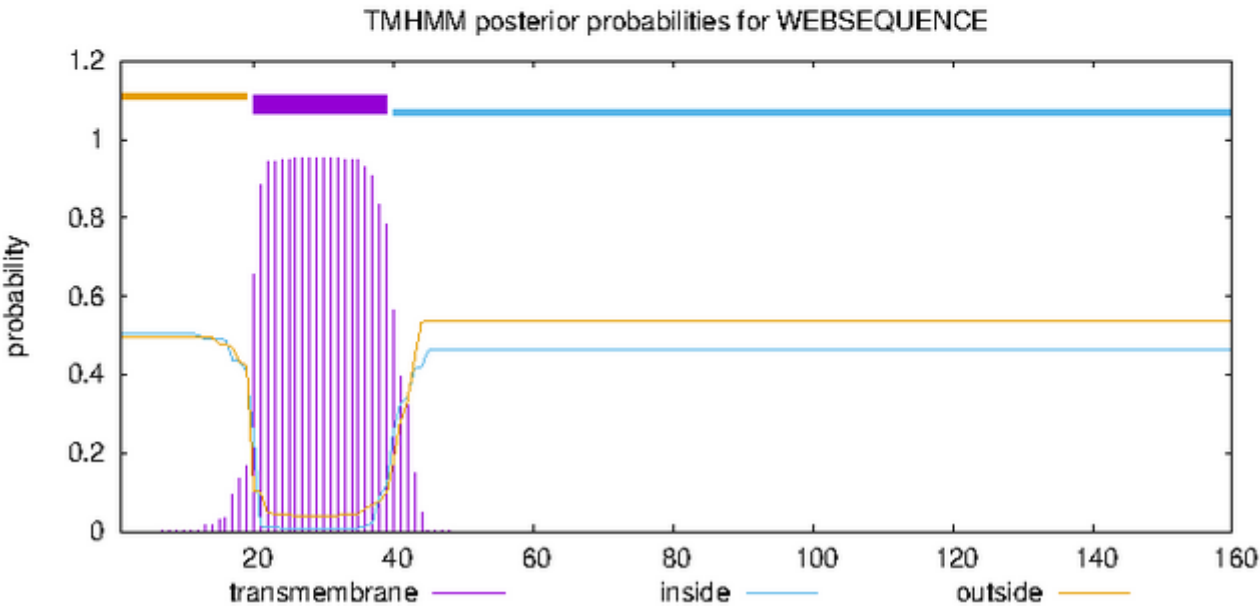
Targetname/Identifier: fl690RF\_0459919

TMHMM result

<https://services.healthtech.dtu.dk/services/TMHMM-2.0/>

**TMHMM result**

```
# WEBSEQUENCE Length: 160
# WEBSEQUENCE Number of predicted TMHs: 1
# WEBSEQUENCE Exp number of AAs in TMHs: 20.26437
# WEBSEQUENCE Exp number, first 60 AAs: 20.26419
# WEBSEQUENCE Total prob of N-in: 0.50439
# WEBSEQUENCE POSSIBLE N-term signal sequence
WEBSEQUENCE TMHMM2.0 outside 1 19
WEBSEQUENCE TMHMM2.0 TMhelix 20 39
WEBSEQUENCE TMHMM2.0 inside 40 160
```



Targetname/Identifier: fl690RF\_0459919

Potential Cleavage Sites

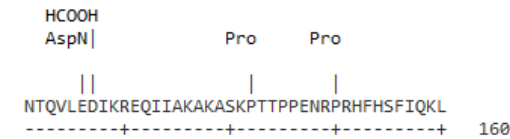
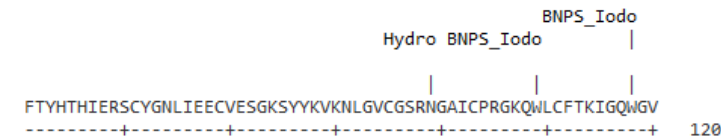
[https://web.expasy.org/peptide\\_cutter/](https://web.expasy.org/peptide_cutter/)

These enzymes cleave the sequence:

Name of enzyme	No. of cleavages	Positions of cleavage sites
Asp-N endopeptidase	1	126
BNPS-Skatole	2	109 118
CNBr	3	1 20 57
Formic acid	1	127
Hydroxylamine	1	99
Iodosobenzoic acid	2	109 118
Proline-endopeptidase [*]	2	142 150

These are the cleavage sites of the chosen enzymes and chemicals mapped onto the entered protein sequence:

- You have chosen a block size of **60** for the map.
- Please note that the cleavage occurs at the **right side** (C-terminal direction) of the marked amino acid.
- You have the possibility to display the results of a single enzyme by **mouseclicking** on the respective enzyme name in the map.





Targetname/Identifier: fl69ORF\_0459919

**O-GalNAc (Mucin Type) Glycosylation Sites in mammalian proteins**

<https://services.healthtech.dtu.dk/services/NetOGlyc-4.0/>

##gff-version 2  
##source-version NetOGlyc 4.0.0.13  
##date 24-4-29  
##Type Protein

#seqname	source	feature	start	end	score	strand	frame	comment		
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		7		7		0.419568	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		8		8		0.200149	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		11		11		0.336192	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		12		12		0.363725	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		15		15		0.184677	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		17		17		0.185744	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		23		23		0.0452218	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		25		25		0.0113495	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		26		26		0.00786615	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		31		31		0.0510987	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		37		37		0.157973	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		38		38		0.202461	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		43		43		0.0276734	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		50		50		0.0198441	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		62		62		0.0164751	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		65		65		0.00429985	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		70		70		0.00637665	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		82		82		1.15605e-05	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		85		85		0.00388972	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		97		97		0.027459	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		113		113		0.0116779	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		122		122		0.0101992	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		140		140		0.295608	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		143		143		0.436329	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		144		144		0.37979	.	.
SEQUENCE	netOGlyc-4.0.0.13	CARBOHYD		155		155		0.0633676	.	.

Targetname/Identifier: fl690RF\_0459919

**N-linked glycosylation sites in human proteins**  
<https://services.healthtech.dtu.dk/services/NetNGlyc-1.0/>

**Output for 'Sequence'**

```
#####

Warning: This sequence may not contain a signal peptide!!

Proteins without signal peptides are unlikely to be exposed to
the N-glycosylation machinery and thus may not be glycosylated
(in vivo) even though they contain potential motifs.

SignalP-NN euk predictions are as follows:

# name          Cmax  pos ?  Ymax  pos ?  Smax  pos ?  Smean ?  D      ?
Sequence        0.299  40  0.437  40  0.838  33  0.374  0.412 N  0.500      SignalP-TM

SignalP output is explained at https://services.healthtech.dtu.dk/services/SignalP-4.1/output.php

#####

Name: Sequence          Length: 160
MACIYPTTFYTSLPTKSLNMGISLTTILILSVAVLLSTAAPPSCRECYQSLHYRGEMQQYFTYHTHIERSCYGNLIEECV      80
ESGKSYKVKNLGVCGRNGAICPRGKQWLCTKIGQWGVNTQVLEDIKREQIIAKAKASKPTTPPENRPRHFHSFIQKL      160
.....                      80
.....                      160

(Threshold=0.5)

No sites predicted in this sequence.
```



Targetname/Identifier: fl690RF\_0459919

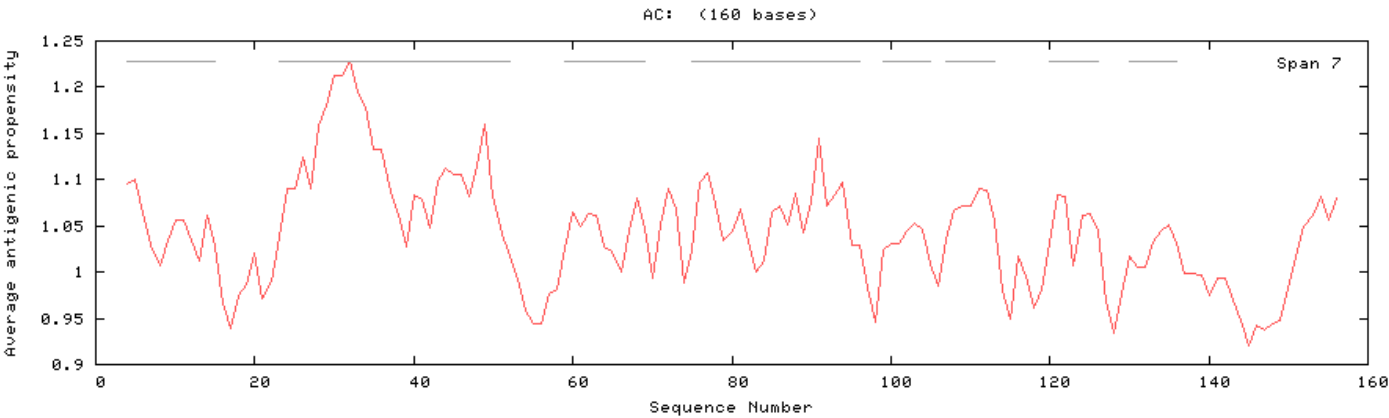
Antigenic Plot

<http://imed.med.ucm.es/Tools/antigenic.pl>

Your sequence is 160 residues long

Average antigenic propensity for this protein is 1.0441

Antigenic plot for sequence



There are 8 antigenic determinants in your sequence:

n	Start Position	Sequence	End Position
1	4	IYPTTFYTSLPT	15
2	23	SLTTILILSVAVLLSTAAPPSCRECYQSLH	52
3	59	QYFTYHTHIER	69
4	75	LIEECVESGKSYYKVKNLGVCG	96
5	99	NGAICPR	105
6	107	KQWLCFT	113
7	120	VNTQVLE	126
8	130	REQUIAK	136

Contact **Pedro Reche**  
Last Update: 29 April 2024

Targetname/Identifier: xxxxxx

MPI Toolkit – Quick2D  
<https://toolkit.tuebingen.mpg.de/tools/quick2d>

Protein ID: Q\_7360255

We have detected a potential signal peptide in your query protein!

