A brief introduction to ExPASy

ExPASy网站应用实例

Group 05

25/01/2015

- EXPASY是由瑞士生物信息研究所（SIB）维护的蛋白组学分析平台，侧重于蛋白序列，结构及2-D 电泳数据的分析。
- 整合了很多蛋白质数据资源和分析工具
- 由实验生物学家参与数据库的构建，数据注释质量高

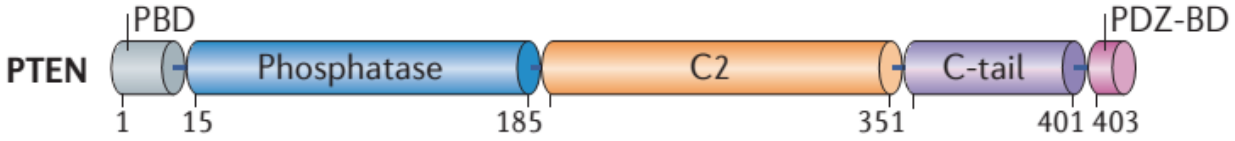
Contents:

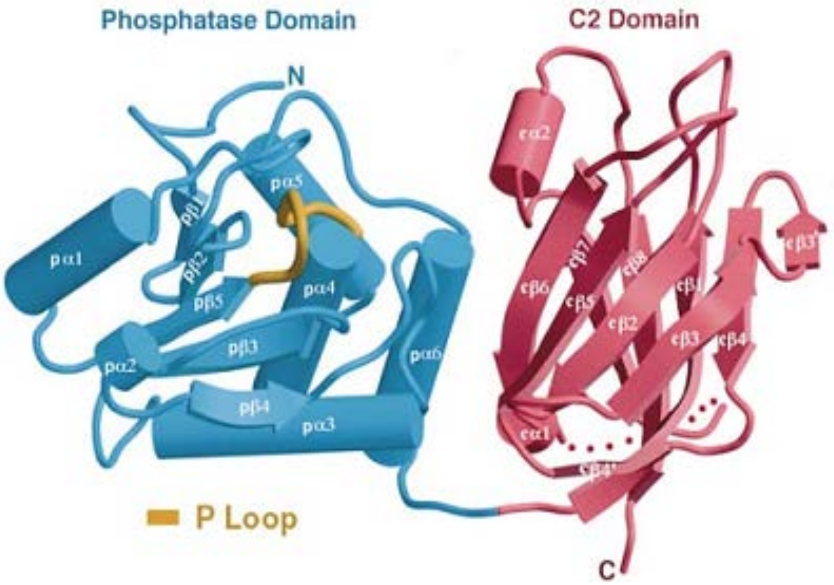
Function of PTEN

- PTEN protein is found in almost all tissues in the body.
- PTEN is an antagonist of PI3K signaling.
- PTEN is regulated by PTMs that affect its localization, stability, and activity.
- PTEN-Long variant was predicted to contain an additional 173 amino acids at the amino terminus of the canonical PTEN protein

Hopkins BD et al. ***Trends Biochem Sci.*** 2014 Apr;39(4):183-90.
Seton-Rogers S. ***Nat Rev Cancer.*** 2013 Aug;13(8):520.

Structure of PTEN tumor suppressor





Song MS, et al. ***Nat Rev Mol Cell Biol.*** 2012 Apr 4;13(5):283-96.
Lee JO, et al. ***Cell.*** 1999 Oct 29;99(3):323-34.

Protein sequences and identification

- pI, Mw 等电点，分子量
- AA 氨基酸组成
- Hydrophobicity /hydrophilicity 亲疏水性
- Enzymatic cleavage site 酶切位点
- Extinction Coefficient 消光系数

Protein sequences and identification

Tools	Functions
AACompIdent	Protein identification by aa composition
Compute pI/MW	Theoretical pI and Mw computation
ProtParam	Protein physical and chemical parameters
ProtScale	Protein profile computation and representation
HAMAP	UniProtKB family classification and annotation
LALIGN	Pairwise alignment
PeptideMass	Peptides from protein cleavage
TagIdent	Protein identification with PI, Mw

Prediction of PTEN' molecular weight and pI by Cmpute pI/Mw

Compute pI/Mw

Theoretical pI/Mw (average) for the user-entered sequence:

```
        10         20         30         40         50         60
MTAIIKEIVS RNKRRYQEDG FDLDLTYIYP NIIAMGFPAE RLEGVYRNNI DDVVRFLDSK

        70         80         90        100        110        120
HKNHYKIYNL CAERHYDTAK FNCRVAQYPF EDHNPPQLEL IKPFCEDLDQ WLSEDDNHVA

       130        140        150        160        170        180
AIHCKAGKGR TGVMICAYLL HRGKFLKAQE ALDFYGEVRT RDKKGVTIPS QRRYVYYYSY

       190        200        210        220        230        240
LLKNHLDYRP VALLFHKMMF ETIPMFSGGT CNPQFVVCQL KVKIYSSNSG PTRREDKFMY

       250        260        270        280        290        300
FEFPQPLPVC GDIKVEFFHK QNKMLKKDKM FHFWVNTFFI PGPEETSEKV ENGSLCDQEI

       310        320        330        340        350        360
DSICSIERAD NDKEYLVLTL TKNDLDKANK DKANRYFSPN FKVKLYFTKT VEEPSNPEAS

       370        380        390        400
SSTSVTPDVS DNEPDHYRYS DTTDSDPENE PFDEDQHTQI TKV
```

Theoretical pI/Mw: 5.94 / 47166.29

ProtParam tool: Another tool for AA Composition, pI and Mw

Number of amino acids: 403

Molecular weight: 47166.2

Theoretical pI: 5.94

Amino acid composition:

Ala (A)	17	4.2%
Arg (R)	20	5.0%
Asn (N)	23	5.7%
Asp (D)	33	8.2%
Cys (C)	10	2.5%
Gln (Q)	13	3.2%
Glu (E)	29	7.2%
Gly (G)	15	3.7%
His (H)	13	3.2%
Ile (I)	20	5.0%
Leu (L)	27	6.7%
Lys (K)	34	8.4%
Met (M)	9	2.2%
Phe (F)	25	6.2%
Pro (P)	23	5.7%
Ser (S)	22	5.5%
Thr (T)	21	5.2%
Trp (W)	2	0.5%
Tyr (Y)	23	5.7%
Val (V)	24	6.0%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 62
Total number of positively charged residues (Arg + Lys): 54

Atomic composition:

Carbon	C	2113
Hydrogen	H	3215
Nitrogen	N	561
Oxygen	O	630
Sulfur	S	19

Formula: $C_{2113}H_{3215}N_{561}O_{630}S_{19}$
Total number of atoms: 6538

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 45895
Abs 0.1% (=1 g/l) 0.973, assuming all pairs of Cys residues form cystines

Ext. coefficient 45270
Abs 0.1% (=1 g/l) 0.960, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

ProtScale: a numerical value assigned to each type of amino acid.

ProtScale [Reference / Documentation] allows you to compute and represent the profile produced by any amino acid scale on a selected protein.

An **amino acid scale** is defined by a numerical value assigned to each type of amino acid. The most frequently used scales are the hydrophobicity or hydrophilicity scales and the secondary structure conformational parameters scales, but many other scales exist which are based on different chemical and physical properties of the amino acids. This program provides 57 predefined scales entered from the literature.

Enter a UniProtKB/Swiss-Prot or UniProtKB/TrEMBL accession number (AC) (e.g. **P05130**) or a sequence identifier (ID) (e.g. **KPC1_DROME**):

Or you can paste your own sequence in the box below:

Please choose an amino acid scale from the following list. To display information about a scale (author, reference, amino acid scale values) you can click on its name.

- ○ Molecular weight
- ○ Bulkiness
- ○ Polarity / Grantham
- ○ Recognition factors
- ○ Hphob. OMH / Sweet et al.
- ◉ Hphob. / Kyte & Doolittle
- ○ Hphob. / Abraham & Leo
- ○ Hphob. / Bull & Breese
- ○ Hphob. / Guy
- ○ Hphob. / Miyazawa et al.
- ○ Hphob. / Roseman
- ○ Hphob. / Wolfenden et al.
- ○ Number of codon(s)
- ○ Polarity / Zimmerman
- ○ Refractivity
- ○ Hphob. / Eisenberg et al.
- ○ Hphob. / Hopp & Woods
- ○ Hphob. / Manavalan et al.
- ○ Hphob. / Black
- ○ Hphob. / Fauchere et al.
- ○ Hphob. / Janin
- ○ Hphob. / Rao & Argos
- ○ Hphob. / Tanford
- ○ Hphob. / Welling & al

ProtScale: a numerical value assigned to each type of amino acid.

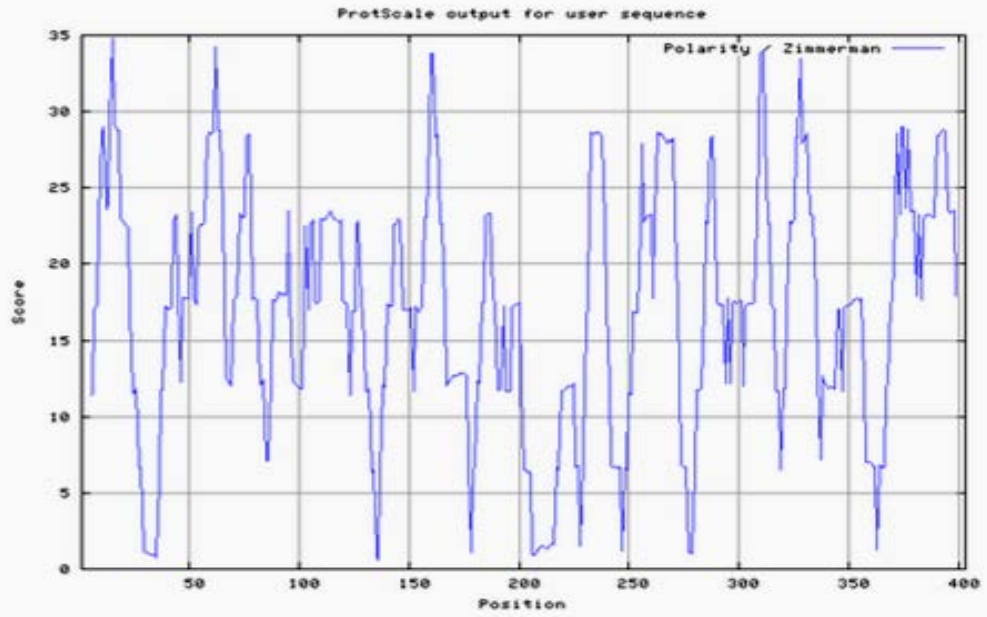
Using the scale **Polarity / Zimmerman**, the individual values for the 20 amino acids are:

Ala:	0.000	Arg:	52.000	Asn:	3.380	Asp:	49.700	Cys:	1.480
Gln:	3.530	Glu:	49.900	Gly:	0.000	His:	51.600	Ile:	0.130
Leu:	0.130	Lys:	49.500	Met:	1.430	Phe:	0.350	Pro:	1.580
Ser:	1.670	Thr:	1.660	Trp:	2.100	Tyr:	1.610	Val:	0.130

: 26.540 : 26.715 : 13.594

Weights for window positions 1,..,9, using **exponential weight variation model**:

1	2	3	4	5	6	7	8	9
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
edge				center				edge



PeptideMass: cleaves a protein sequence from a user-entered protein sequence with a chosen enzyme, and computes the masses of the generated peptides.

PeptideMass

The entered sequence is:

```
        10         20         30         40         50         60
MTAIIKEIVS RNKRRYQEDG FDLDLTYIYP NIIAMGFPAE RLEGVYRNNI DDVVRFLDSK

        70         80         90        100        110        120
HKNHYKIYNL CAERHYDTAK FNCRVAQYPF EDHNPPQLEL IKPFCEDLDQ WLSEDDNHVA

       130        140        150        160        170        180
AIHCKAGKGR TGVMICAYLL HRGKFLKAQE ALDFYGEVRT RDKKGVTIPS QRRYVYYYSY

       190        200        210        220        230        240
LLKNHLDYRP VALLFHKMMF ETIPMFSGGT CNPQFVVCQL KVKIYSSNSG PTRREDKFMY

       250        260        270        280        290        300
FEFPQPLPVC GDIKVEFFHK QNKMLKKDKM FHFWVNTFFI PGPEETSEKV ENGSLCDQEI

       310        320        330        340        350        360
DSICSIERAD NDKEYLVLTL TKNDLDKANK DKANRYFSPN FKVKLYFTKT VEEPSNPEAS

       370        380        390        400
SSTSVTPDVS DNEPDHYRYS DTTDSDPENE PFDEDQHTQI TKV
```

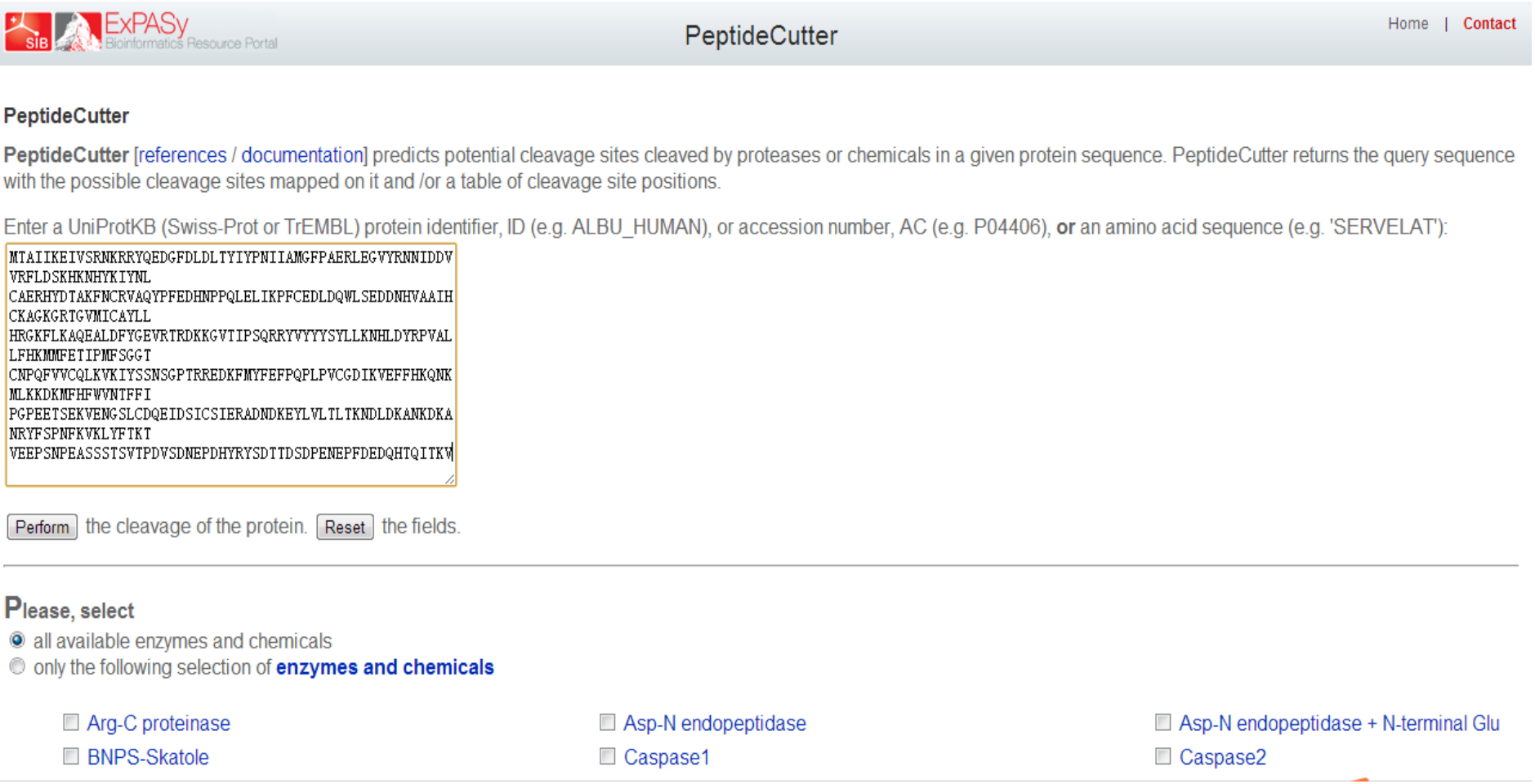
The selected enzyme is: Trypsin

'he peptide masses from your sequence are:

[Theoretical pI: 5.94 / Mw (average mass): 47166.29 / Mw (monoisotopic mass): 47136.15]

mass	position	#MC	modifications	peptide sequence
4804.2446	85-125	0		VAQYPFEDHNPPQLELIKPF CEDLDQWLSEDDNHVAAIHC K
3146.3668	350-378	0		TVEEPSNPEASSSTSVTPDV SDNEPDHYR
3051.4444	16-41	0		YQEDGFDLDLTYIYPNIIAM GFPAER
2812.1704	379-402	0		YSDTTDSDPENEPFDEDQHT QITK
2708.2413	198-221	0		MMFETIPMFSGGTCNPQFVV CQLK
2443.1427	270-289	0		MFHFWVNTFFIPGPEETSEK
2109.9427	290-308	0		VENGSLCDQEIDSICSIER
2030.9754	238-254	0		FMYFEFPQPLPVCGDIK
1722.9438	184-197	0		NHLDYRPVALLFHK
1397.6695	148-159	0		AQEALDFYGEVR
1376.7177	131-142	0		TGVMICAYLLHR
1374.6980	174-183	0		YVYYYSYLLK
1081.5272	224-233	0		IYSSNSGPTR
1079.6347	314-322	0		EYLVLTLTK
981.4822	67-74	0		IYNLCAER

PeptideCutter

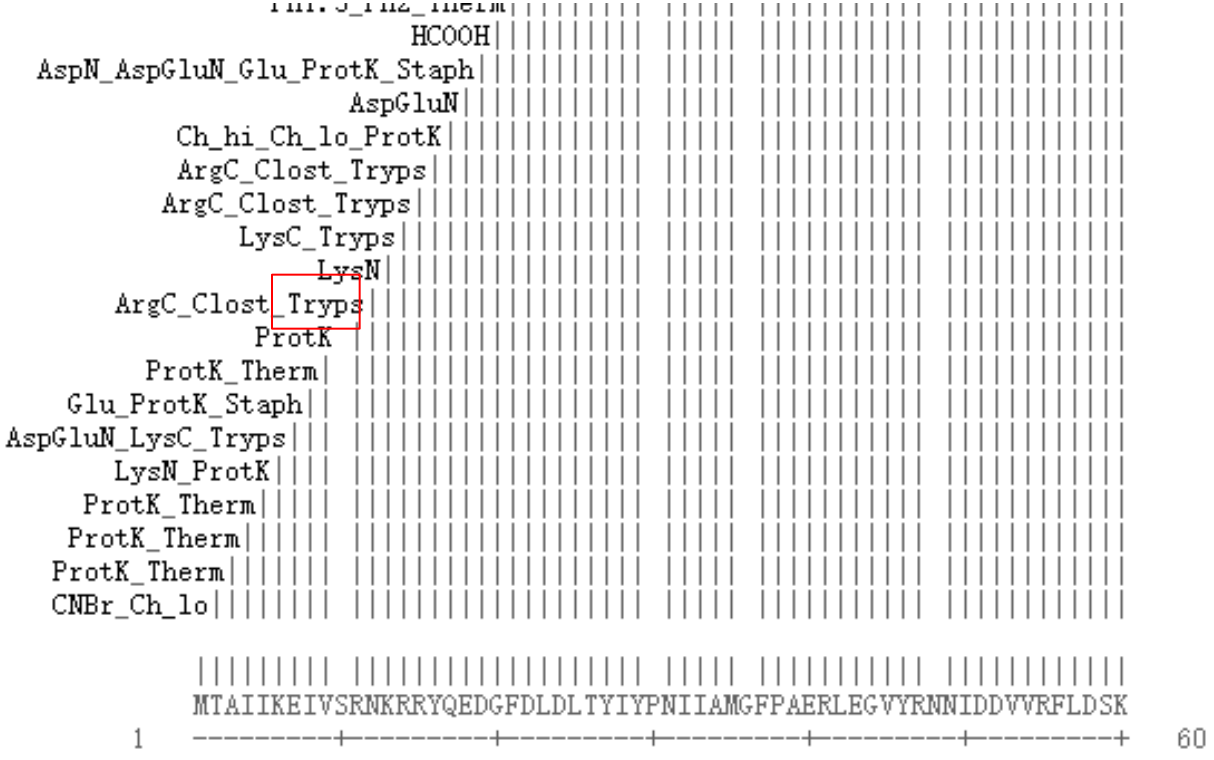


Result of Peptidecutter on PTEN

Name of enzyme	No. of cleavages	Positions of cleavage sites
Arg-C proteinase	20	11 14 15 41 47 55 74 84 130 142 159 161 172 173 189 233 234 308 335 378
Asp-N endopeptidase	33	18 21 23 50 51 57 76 91 106 108 114 115 152 161 186 235 251 267 296 300 309 311 323 325 330 367 370 374 380 383 385 392 394
Asp-N endopeptidase + N-terminal Glu	62	6 17 18 21 23 39 42 50 51 57 72 76 90 91 98 105 106 108 113 114 115 149 152 156 161 186 200 234 235 241 251 255 267 283 284 287 290 296 298 300 306 309 311 313 323 325 330 351 352 357 367 370 372 374 380 383 385 387 389 392 393 394
BNPS-Skatole	2	111 274
CNBr	9	1 35 134 198 199 205 239 264 270
Chymotrypsin-high specificity (C-term to [FYW], not before P)	46	16 21 27 46 56 65 68 76 81 90 104 111 138 145 154 155 174 176 177 178 180 188 195 200 206 215 225 238 240 241 257 258 271 273 274 278 279 315 336 337 341 346 347 377 379 392
Chymotrypsin-low specificity (C-term to [FYWML], not before P)	93	1 16 21 23 25 27 35 42 46 56 57 61 64 65 68 70 75 76 81 90 93 98 100 104 108 111 112 118 123 134 138 139 140 141 145 146 152 154 155 174 176 177 178 180 181 182 185 186 188 193 194 195 196 198 199 200 205 206 215 220 225 238 240 241 257 258 259 264 265 270 271 272 273 274 278 279 295 315 316 318 320 325 336 337 341 345 346 347 376 377 379 392 397
Clostripain	20	11 14 15 41 47 55 74 84 130 142 159 161 172 173 189 233 234 308 335 378
Formic acid	33	19 22 24 51 52 58 77 92 107 109 115 116 153 162 187 236 252 268 297 301 310 312 324 326 331 368 371 375 381 384 386 393 395
Glutamyl endopeptidase	29	7 18 40 43 73 91 99 106 114 150 157 201 235 242 256 284 285 288 291 299 307 314 352 353 358 373 388 390 394

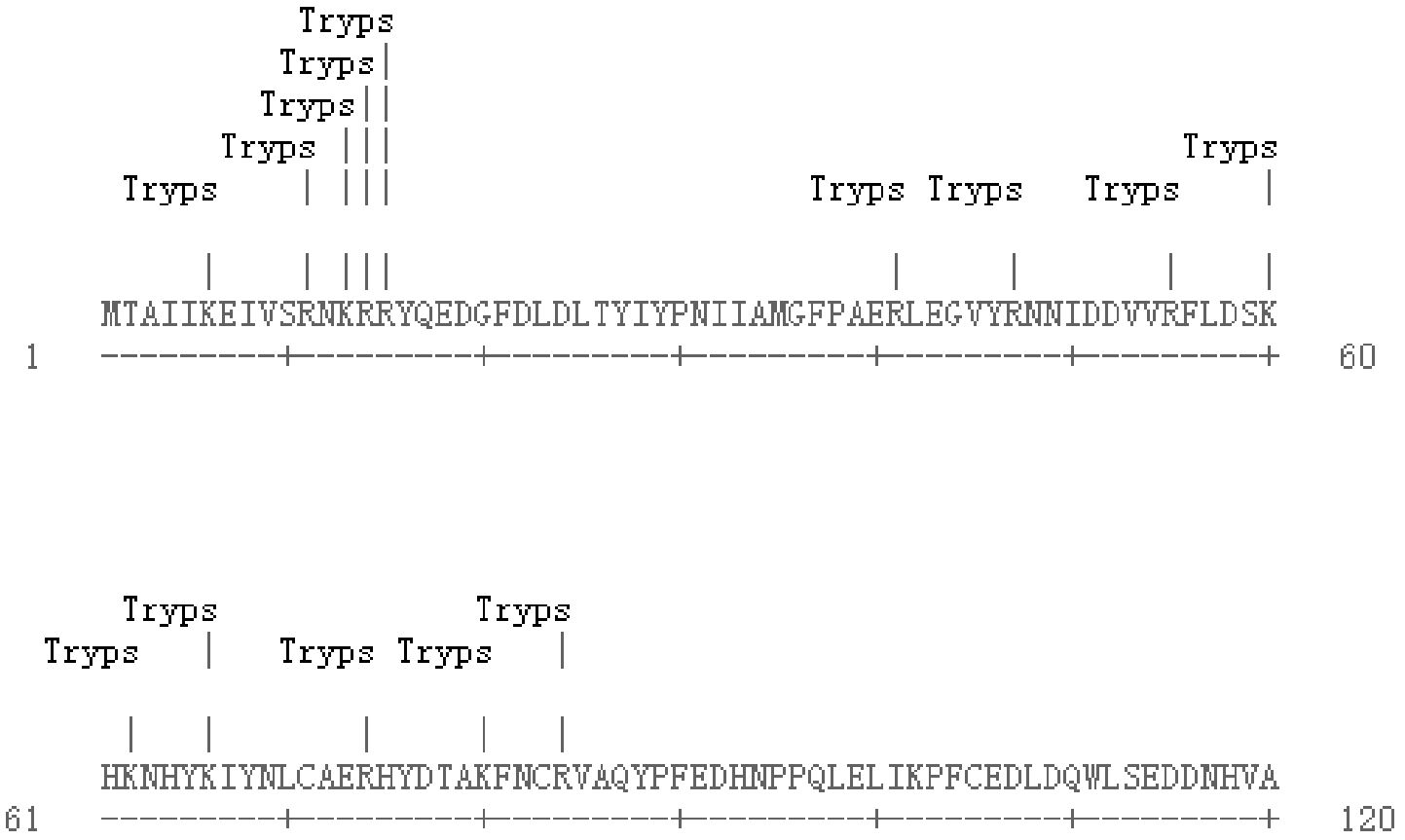
These chosen enzymes do not cut:

Caspase1
Caspase10
Caspase2
Caspase3
Caspase4
Caspase5
Caspase6
Caspase7
Caspase8
Caspase9
Enterokinase
Factor Xa
GranzymeB
Thrombin
Tobacco etch virus prote



These enzymes cleave the sequence:

Name of enzyme	No. of cleavages	Positions of cleavage sites
Trypsin	52	6 11 13 14 15 41 47 55 60 62 66 74 80 84 125 128 130 142 144 147 159 161 163 164 172 173 183 197 221 223 233 234 237 254 260 263 266 267 269 289 308 313 322 327 330 332 335 342 344 349 378 402



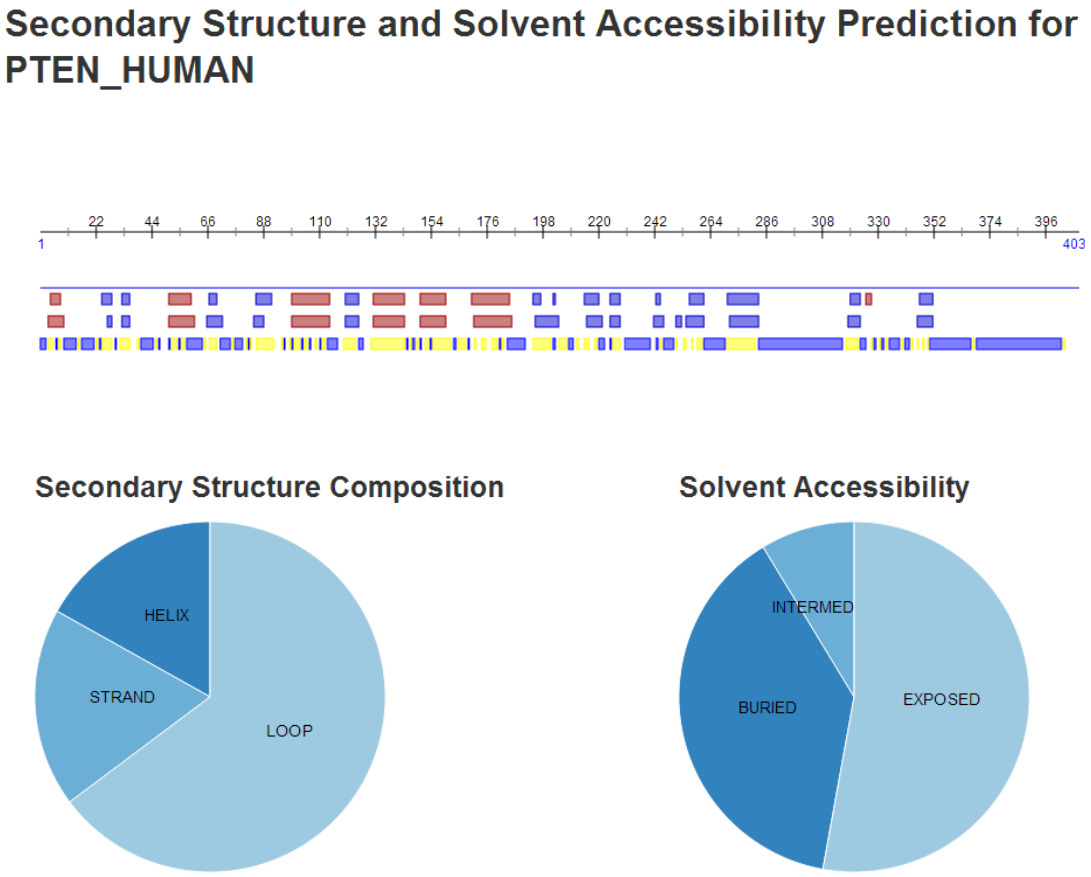
Protein structure prediction

- Secondary structure
- 3D structure
- Domain

Protein structure prediction

Tools	*Functions*
CFSSP	*Protein secondary structure prediction*
COILS	*Prediction of Coiled Coil Regions in Proteins*
DisEMBL	*Prediction of disordered protein regions*
Predict Protein	*Prediction of physico-chemical protein properties*
PORTER	*Protein Secondary Structure Prediction*
PROSITE	*Protein domains and families*
Phyre2	*3D structure prediction with HMM*
SWISS MODEL	*Ptructure homology-modeling*
SWISS pdb viwer	*Analyse protein 3D structures*

The secondary structure of PTEN



Protein Disorder and Flexibility Prediction for PTEN_HUMAN

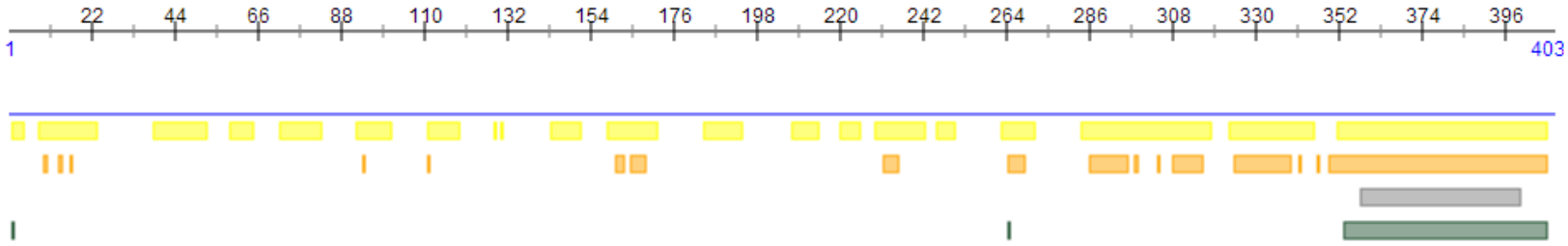
Export

Visual Help

What am I seeing Here? This viewer lays out predicted features that correspond to regions within the queried sequence. Mouse over the different colored boxes to learn more about the annotations

Zoom - Start:1, End:403

Export to image



Disulphide Bridges Prediction for PTEN_HUMAN

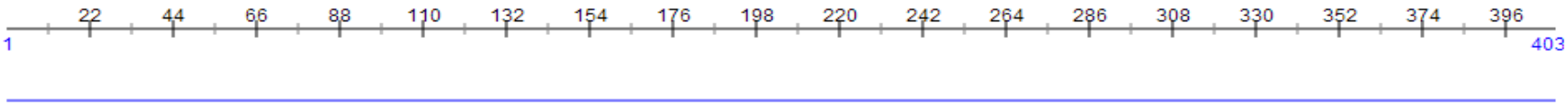
Export

Visual Help

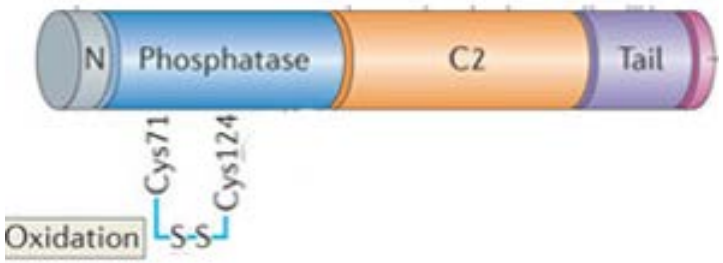
What am I seeing Here? This viewer lays out predicted features that correspond to regions within the queried sequence. Mouse over the different colored boxes to learn more about the annotations

Zoom - Start:1, End:403

Export to image



No disulphide was predicted by Disulphide Bridge Prediction



The Cys124 residue in the active site of PTEN forms a disulfide bond with Cys71

Lee SR et al. ***J Biol Chem.*** 2002 Jun 7;277(23):20336-42.
Song MS, et al. ***Nat Rev Mol Cell Biol.*** 2012 Apr 4;13(5):283-96.

Protein domains prediction of PTEN

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 (release 20.110) motifs on sequence USERSEQ1 :

found: 3 hits in 1 sequence

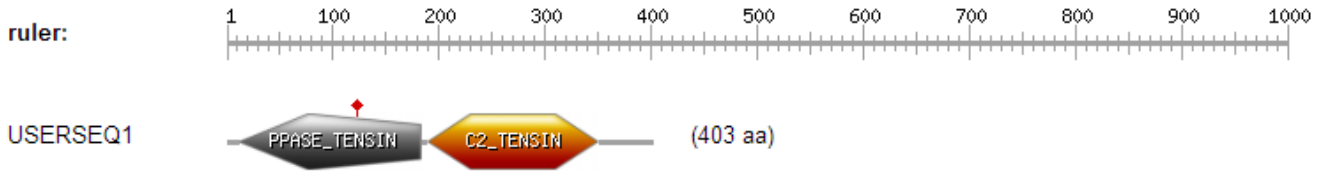
USERSEQ1 (403 aa)

MTAIIKEIVSRNKRRYQEDGFDLDLTYIYPNIIAMGFPAERLEGVYRNNIDDVVRFLDSKHKNHYK
IYNLCAERHYDTAKFNCRVAQYPFEDHNPPQLELIKPFCEDLDQWLSEDDNHVAAIHCKAGKGRTG
VMICAYLLHRGKFLKAQEALDFYGEVRTRDKKGVTIPSQRRYVYYYSYLLKNHLDYRPVALLFHKM
MFETIPMFSGGTCNPQFVVCQLKVKIYSSNSGPTRREDKFMYFEFPQPLPVCGDIKVEFFHKQNKM
LKKDKMFHFWVNTFFIPGPEETSEKVENGSLCDQEIDSICSIERADNDKEYLVLTLTKNDLDKANK
DKANRYFSPNFKVKLYFTKTVEEPSNPEASSSTSVTPDVSDNEPDHYRYSDTTDSDPENEPFDEDQ
HTQITKV

Protein domains prediction of PTEN

hits by profiles: [2 hits (by 2 distinct profiles) on 1 sequence]

Upper case represents match positions, lower case insert positions, and the '-' symbol represents deletions relative to the matching profile.



PS51181 **PPASE_TENSIN** *Phosphatase tensin-type domain profile* :

14 - 185: **score** = 37.280

RRYQEDGFDLDLTYIYPNIIAMGFPAERLEGVYRNNIDDVVRFLDSKHKNHYKIYNLC-A
ERHYDTAKFNCRVAQYPFEDHNPPQLELIKPFCEDLDQWLSEDDNHVAAIHCKAGKGRTG
VMICAYLLHRGKFLKAQEALDFYGEVRTRDK--KGVTIPSQRRYVYYYSYLLKNH

Predicted features:

DOMAIN	14	185	Phosphatase tensin-type	[condition: none]
ACT_SITE	124		Phosphocysteine intermediate	[condition: C]

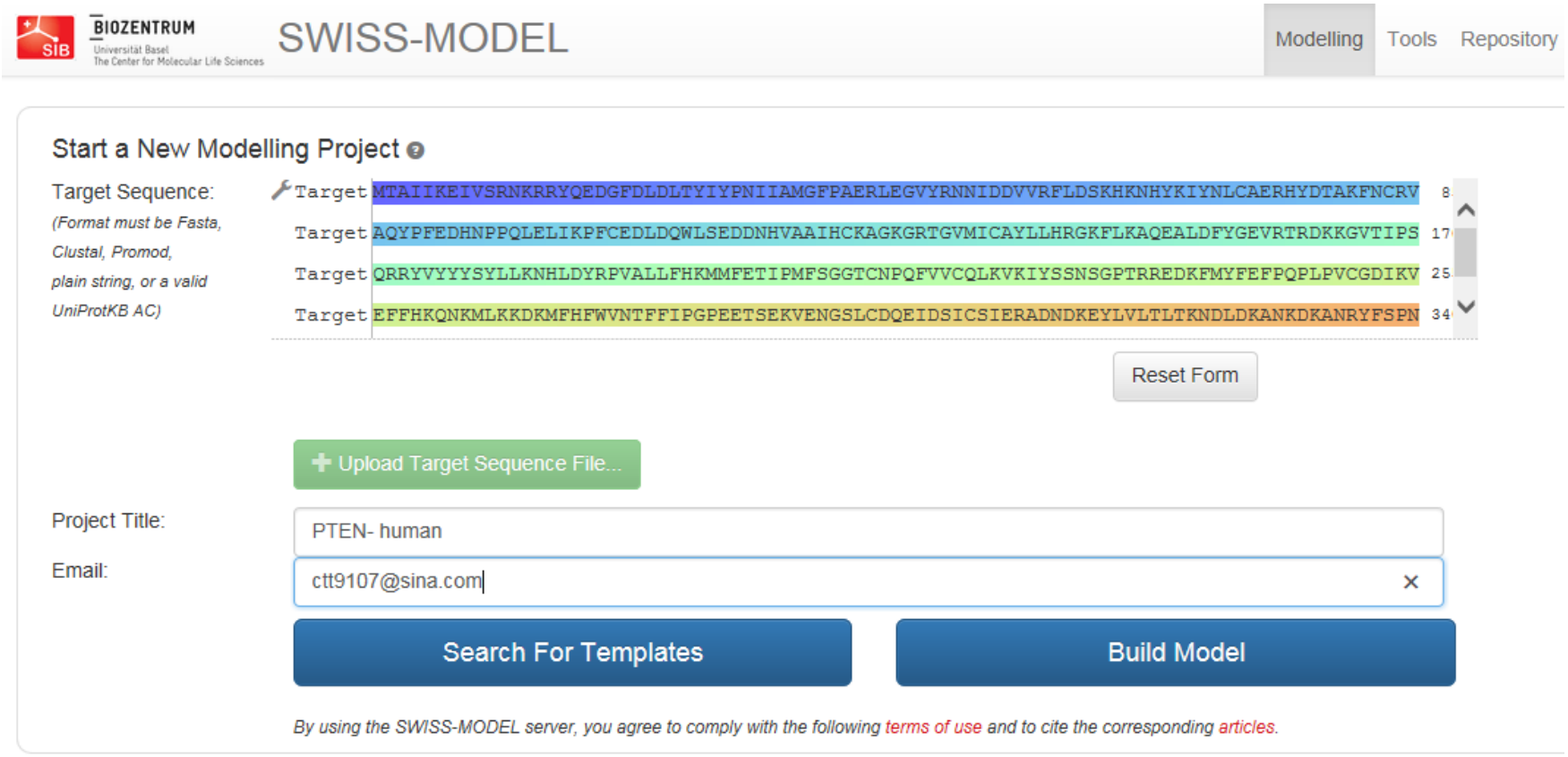
PS51182 **C2_TENSIN** *C2 tensin-type domain profile* :

190 - 350: **score** = 24.203

PVALLFHKMMFETIPMFSGGT-CNPQFVVCQLKVKIYSSNSGPTRR---EDKFMYFEFPQ
PLPVCGDIKVEFFHKQNKMLKKDKMFHFWVNTFFIPGPeetsekvengslcdqeidsics
ieradndkeylVLTLTKNDLDKANkDKANRyFSPNFKVKLYFTKT

3D structure prediction of PTEN

- Swiss-model

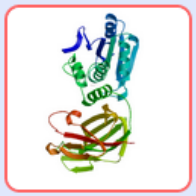


PTEN Created: today at 13:47

Summary | Templates 50 | Models 3

Model Results

Order by: GMQE



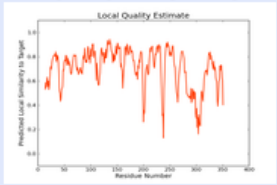
Model 01

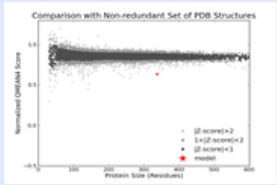
Oligo-State	Ligands	GMQE	QMEAN4
MONOMER	1 x TLA	0.73	-6.09

1 x L(+)-TARTARIC ACID

Ligand 1 in contact with: Chain A : D92, H93, C124, K125, A126, G129, R130, Q171

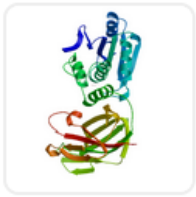
QMEAN4	-6.09
Cβ	-4.20
All Atom	-3.14
Solvation	-4.09
Torsion	-3.61





Template	Seq Identity	Coverage	Description
1d5r.1.A	99.69%		PHOSPHOINOSITIDE PHOSPHATASE PTEN

Model-Template Alignment



Model 02

Oligo-State	Ligands	GMQE	QMEAN4
MONOMER	1 x TLA	0.68	-6.09

Template	Seq Identity	Coverage	Description
1d5r.1.A	99.38%		PHOSPHOINOSITIDE PHOSPHATASE PTEN

Model-Template Alignment

Model 03

Oligo-State	Ligands	GMQE	QMEAN4
MONOMER	None	0.55	-10.99

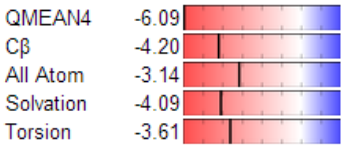
Template	Seq Identity	Coverage	Description
3v0g.2.A	35.53%		Voltage-sensor containing phosphatase

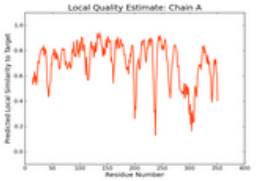
Model-Template Alignment

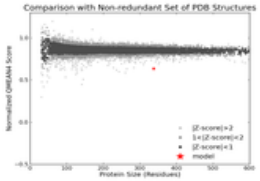
Models

The following model was built (see Materials and Methods "Model Building"):

Model #01	File	Built with	Oligo-State	Ligands	GMQE	QMEAN4
	PDB	ProMod Version 3.70.	MONOMER	1 x TLA: L(+)-TARTARIC ACID;	0.73	-6.09







Template	Seq Identity	Oligo-state	Found by	Method	Resolution	Seq Similarity	Range	Coverage	Description
1d5r.1.A	99.69	monomer	BLAST	X-ray	2.10Å	0.62	14 - 351	0.80	PHOSPHOINOSITIDE PHOSPHATASE PTEN

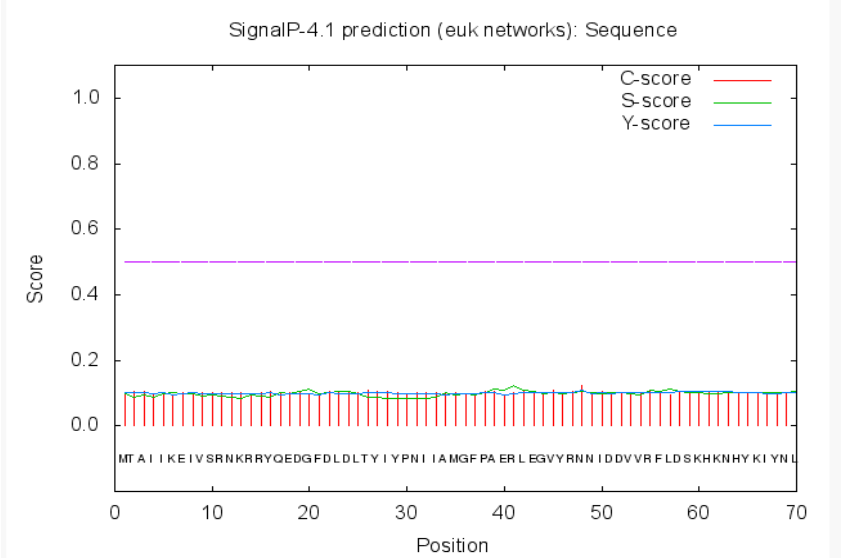
Ligand	Added to Model	Description
TLA	✓	L(+)-TARTARIC ACID

```
Target    MTAIIKEIVSRNKRRYQEDGFDLDLTYIYPNIIAMGFPAERLEGVYRNNIDDVVRFLDSKHKNHYKIYNLCAERHYDTAK
1d5r.1.A -------IVSRNKRRYQEDGFDLDLTYIYPNIIAMGFPAERLEGVYRNNIDDVVRFLDSKHKNHYKIYNLCAERHYDTAK

Target    FNCRVAQYPFEDHNPPQLELIKPFCEDLDQWLSEDDNHVAAIHCKAGKGRTGVMICAYLLHRGKFLKAQEALDFYGEVRT
1d5r.1.A FNCRVAQYPFEDHNPPQLELIKPFCEDLDQWLSEDDNHVAAIHCKAGKGRTGVMICAYLLHRGKFLKAQEALDFYGEVRT

Target    RDKKGVTIPSQRRYVYYYSYLLKNHLDYRPVALLFHKMMFETIPMFSGGTCNPQFVVCQLKVKIYSSNSGPTRREDKFMY
1d5r.1.A RDKKGVTIPSQRRYVYYYSYLLKNHLDYRPVALLFHKMMFETIPMFSGGTCNPQFVVCQLKVKIYSSNSGPTRREDKFMY

Target    FEFPQPLPVCGDIKVEFFHKQNKMLKKDKMFHFWVNTFFIPGPEETSEKVENGSLCDQEIDSICSIERADNDKEYLVLTL
1d5r.1.A FEFPQPLPVCGDIKVEFFHKQNKMLKKDKMFHFWVNTFFIPGPEE----------------------VDNDKEYLVLTL

Target    TKNDLDKANKDKANRYFSPNFKVKLYFTKTVEEPSNPEASSSTSVTPDVSDNEPDHYRYSDTTDSDPENEPFDEDQHTQI
1d5r.1.A TKNDLDKANKDKANRYFSPNFKVKLYFTKTVEE-----------------------------------------------

Target    TKV
1d5r.1.A ---
```

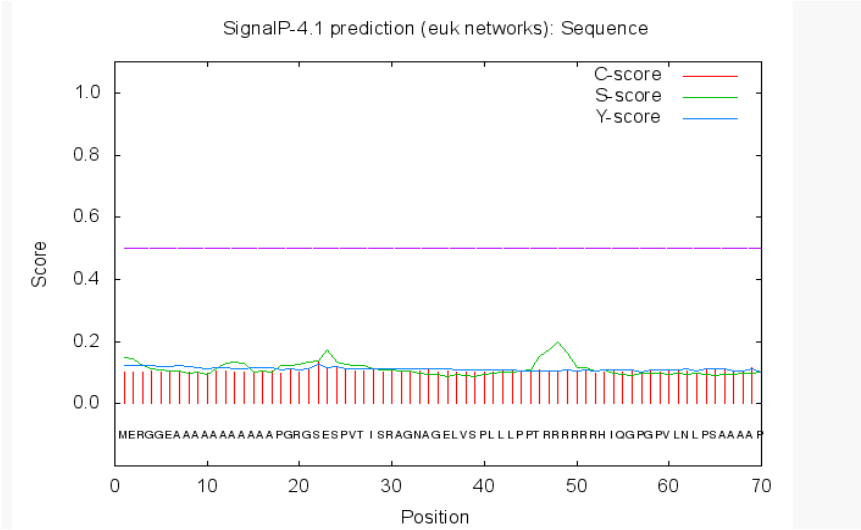
Protein function

- Secretory protein 分泌蛋白
- Transmembrane Protein 膜蛋白
- Subcellular localization 细胞定位
- Post-transcriptional modification 翻译后修饰

Secretory protein: SignalP



PTEN



PTEN-Long

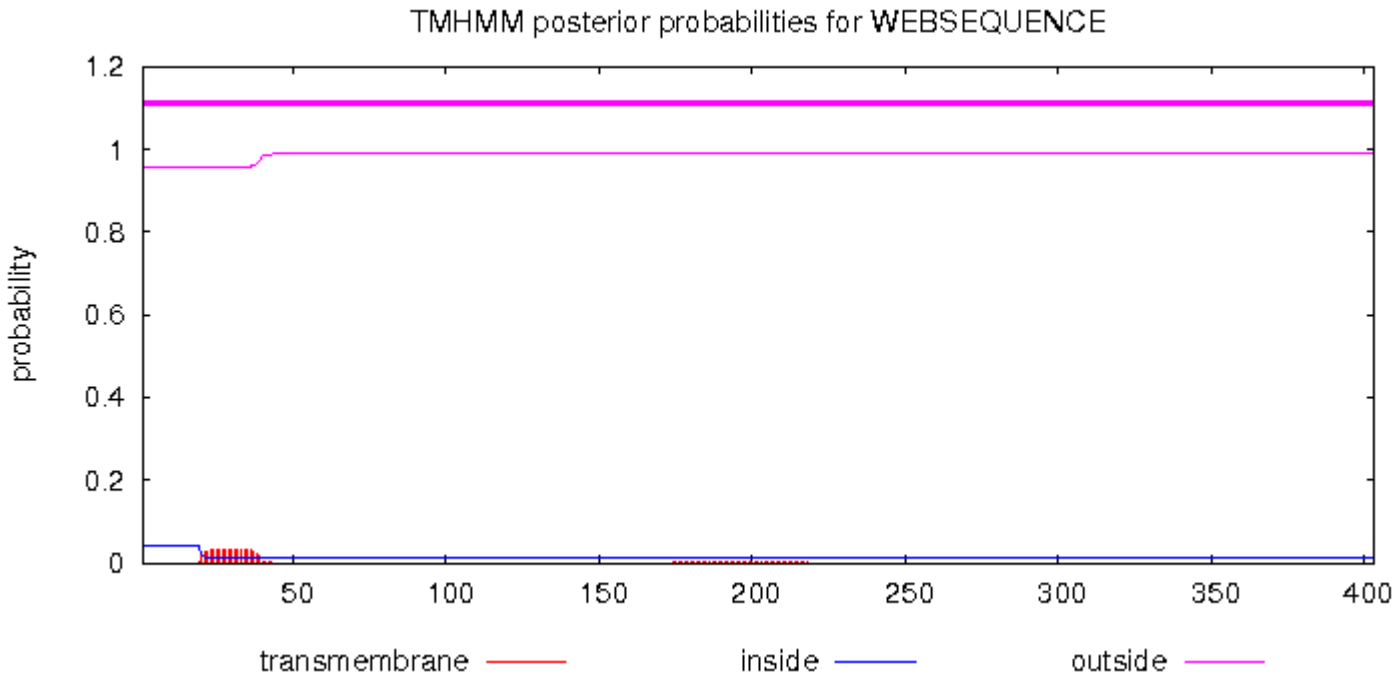
- SignalP indicated that PTEN-Long contained a secretion signal sequence with a predicted cleavage site at amino acid 22.
- Experiment evidence shows that PTEN-Long is secreted from cells and can enter other cells.

Hopkins BD et al. ***Science***. 2013 Jul 26;341(6144):399-402.
Liang H et al. ***Cell Metab.*** 2014 May 6;19(5):836-48.

Prediction of Transmembrane Protein

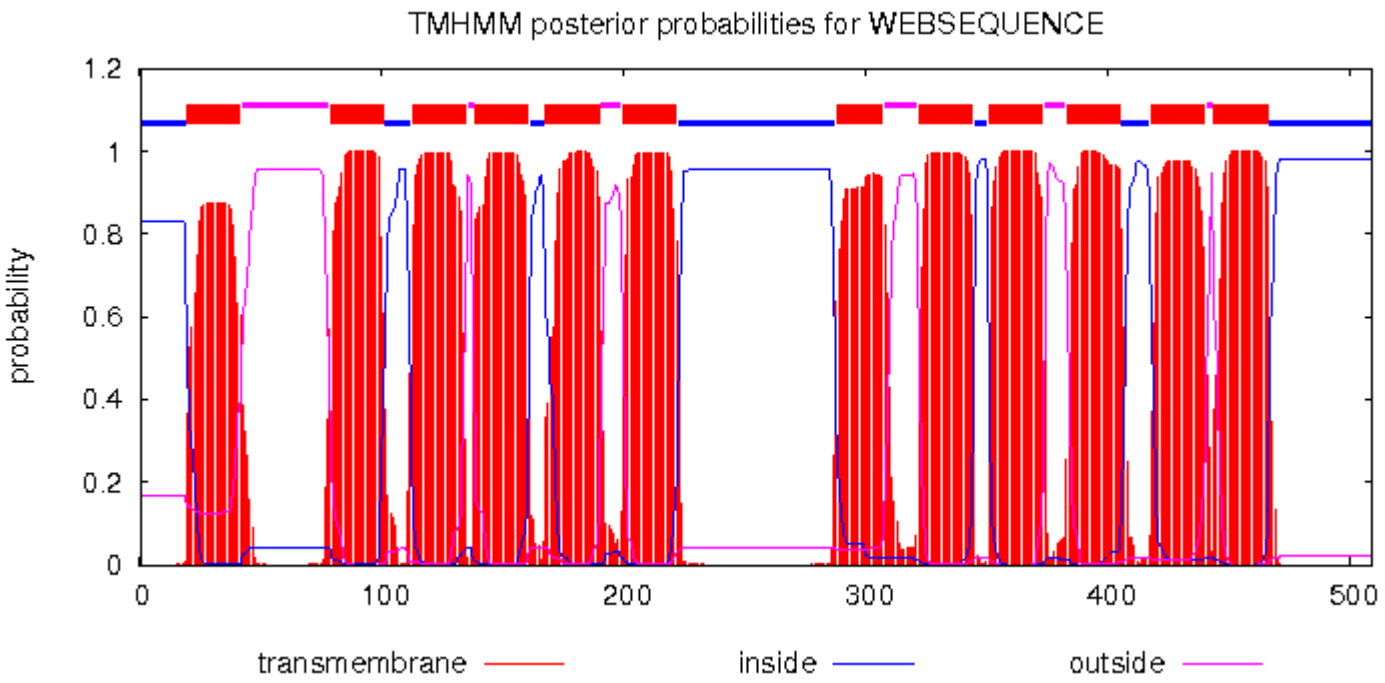
Tools	Functions
DAS-TMfilter	Prediction of transmembrane regions
HMMTOP	Prediction of transmembranes helices and topology
HTMSRAP	Helical TransMembrane Prediction
Phobius	Predict transmembrane topology and signal peptides
TMHMM	Prediction of transmembrane helices in proteins
TMPred	Membrane-spanning region prediction
TopPred	Topology prediction of membrane proteins

TMHMM as an example: prediction of PTEN



PTEN is not a transmembrane protein.

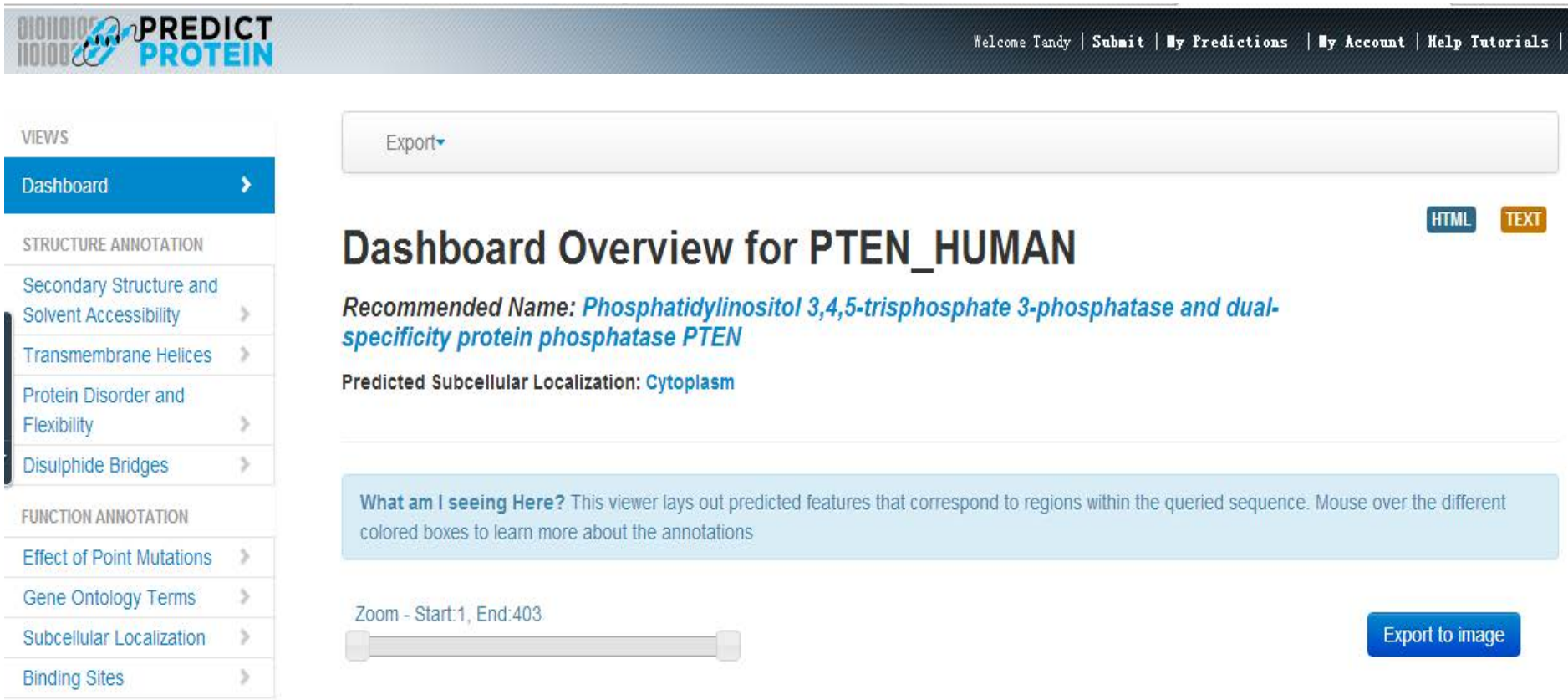
TMHMM as an example: prediction of membrane protein GLUT4



Number of predicted TMHs: 12

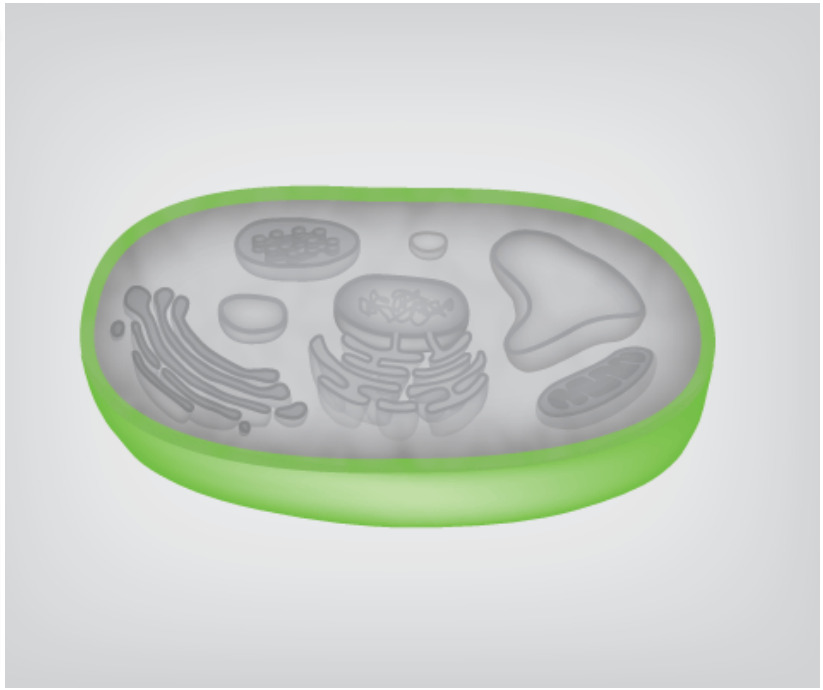
Glucose transporter type 4, also known as GLUT4, is the insulin-regulated glucose transporter found primarily in adipose tissues and striated muscle.

Predict protein:



Subcellular Localization Prediction for PTEN and GLUT4

PTEN



GLUT4

Prediction of Subcellular Localization

Tools	Functions
MITOPROT	Predict mitochondrial targeting sequences
NetNES	Prediction of leucine-rich nuclear export signals
PSORT	Protein subcellular location prediction
TargetP	Subcellular location prediction

Prediction of PTEN's Subcellular Localization

PSORT result:

```
----- Final Results -----

                        cytoplasm --- Certainty= 0.650(Affirmative) < succ>
      mitochondrial matrix space --- Certainty= 0.100(Affirmative) < succ>
                lysosome (lumen) --- Certainty= 0.100(Affirmative) < succ>
endoplasmic reticulum (membrane) --- Certainty= 0.000(Not Clear) < succ>
```

TargetP 1.1 Server - prediction results

```
### targetp v1.1 prediction results ####################################
Number of query sequences:  1
Cleavage site predictions not included.
Using NON-PLANT networks.

Name                    Len             mTP     SP   other  Loc  RC
------------------------------------------------------------------------
Sequence                403           0.155  0.113  0.765   _    2
------------------------------------------------------------------------
cutoff                                0.000  0.000  0.000
```

PTEN is mainly located in the cytoplasm

Prediction of Cytochrome c1's Subcellular Localization

TargetP 1.1 Server - prediction results

```
### targetp v1.1 prediction results ######################################
Number of query sequences:  1
Cleavage site predictions not included.
Using NON-PLANT networks.

Name                    Len            mTP     SP   other  Loc  RC
-----------------------------------------------------------------------
Sequence                548          0.831  0.053  0.095   M    2
-----------------------------------------------------------------------
cutoff                               0.000  0.000  0.000
```

- Cytochrome C1 is targeted to the mitochondrial intermembrane space.
- It is forms the third proton pump in the mitochondrial electron transport chain.

MITOPROT: Predict mitochondrial targeting sequences

Sequence name: PTEN
Input sequence length : 406 aa

VALUES OF COMPUTED PARAMETERS

Net charge of query sequence	: -14
Analysed region	: 6
Number of basic residues in targeting sequence	: 1
Number of acidic residues in targeting sequence	: 0
Cleavage site	**: not predictable**
Cleaved sequence	:

PROBABILITY

of export to mitochondria: **0.3311**

Sequence name: CYTOCHROME C1
Input sequence length : 325 aa

VALUES OF COMPUTED PARAMETERS

Net charge of query sequence	: +8
Analysed region	: 85
Number of basic residues in targeting sequence	: 10
Number of acidic residues in targeting sequence	: 0
Cleavage site	**: 36**
Cleaved sequence	: MAAAAAASLRGVVLGPRGAGLPGARARGLLCSARPG

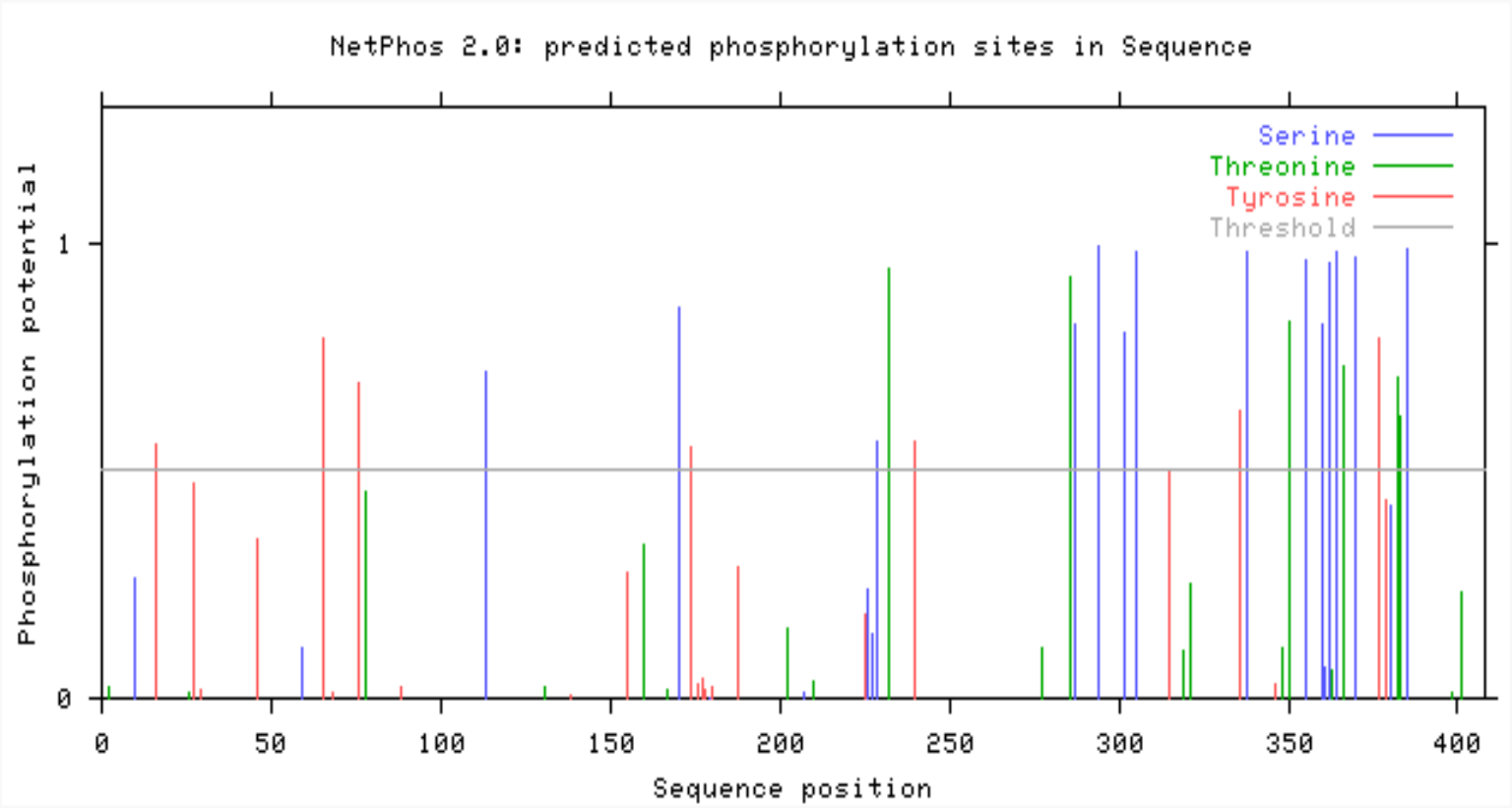
PROBABILITY

of export to mitochondria: **0.8446**

Prediction of Post-transcriptional modification

Tools	Functions
MICSS-Palm	Prediction of palmitoylation sites in proteins
GPS	Prediction of kinase-specific phosphorylation site
NetAcet	predict N-terminal acetylation sites
NetCGlyc	C-mannosylation sites in mammalian proteins
NetCorona	Coronavirus 3C-like proteinase cleavage sites
NetGlycate	glycation of epsilon amino groups of lysines
NetNGlyc	N-glycosylation sites in human proteins
NetOGlyc	mammalian mucin type GalNAc O-glycosylation sites
NetPhos	predict phosphorylation in eukaryotes
NetPhosK	Prediction of kinase-specific phosphorylation site
NMT	Predict N-terminal N-myristoylation of proteins
Sulfinator	tyrosine sulfation site prediction
SulfoSite	Prediction of tyrosine sulfations sites
SUMOplot	Prediction of sumoylation sites in proteins
SUMOsp	Prediction of sumoylation sites
YinOYang	O-beta-GlcNAc attachment sites in eukaryotes

Prediction of phosphorylation of PTEN by NetPhos 2.0



Prediction of phosphorylation of PTEN by NetPhos 2.0

Serine predictions

Name	Pos	Context	Score	Pred
Sequence	10	KEIVSRNKR	0.263	.
Sequence	59	RFLDSKHKN	0.109	.
Sequence	113	DQWLSEDDN	0.720	*S*
Sequence	170	VTIPSQRRY	0.860	*S*
Sequence	179	VYYYSYLLK	0.002	.
Sequence	207	IPMFSGGTC	0.010	.
Sequence	226	VKIYSSNSG	0.239	.
Sequence	227	KIYSSNSGP	0.143	.
Sequence	229	YSSNSGPTR	0.564	*S*
Sequence	287	PEETSEKVE	0.821	*S*
Sequence	294	VENGSLCDQ	0.994	*S*
Sequence	302	QEIDSICSI	0.806	*S*
Sequence	305	DSICSIERA	0.981	*S*
Sequence	338	NRYFSPNFK	0.983	*S*
Sequence	355	VEEPSNPEA	0.963	*S*
Sequence	360	NPEASSSTS	0.822	*S*
Sequence	361	PEASSSTSV	0.068	.
Sequence	362	EASSSTSVT	0.959	*S*
Sequence	364	SSSTSVTPD	0.981	*S*
Sequence	370	TPDVSDNEP	0.966	*S*
Sequence	380	HYRYSDTTD	0.423	.
Sequence	385	DTTDSDPEN	0.989	*S*

SUMOplot: Prediction of Sumoylation sites in proteins

Protein ID:	gi|42560209|sp|P60484.1|PTEN_HUMAN
Defintion:	RecName: Full=Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN; AltName: Full=Mutated in multiple advanced cancers 1; AltName: Full=Phosphatase and tensin homolog
Length:	403 aa

```
  1 MTAIIKEIVS RNKRRYQEDG FDLDLTYIYP NIIAMGFPAE RLEGVYRNNI
 51 DDVVRFLDSK HKNHYKIYNL CAERHYDTAK FNCRVAQYPF EDHNPPQLEL
101 IKPFCEDLDQ WLSEDDNHVA AIHCKAGKGR TGVMICAYLL HRGKFLKAQE
151 ALDFYGEVRT RDKKGVTIPS QRRYVYYYSY LLKNHLDYRP VALLFHKMMF
201 ETIPMFSGGT CNPQFVVCQL KVKIYSSNSG PTRREDKFMY FEFPQPLPVC
251 GDIKVEFFHK QNKMLKKDKM FHFWVNTFFI PGPEETSEKV ENGSLCDQEI
301 DSICSIERAD NDKEYLVLTL TKNDLDKANK DKANRYFSPN FKVKLYFTKT
351 VEEPSNPEAS SSTSVTPDVS DNEPDHYRYS DTTDSDPENE PFDEDQHTQI
401 TKV
```

Motifs with high probability
Motifs with low probability
Overlapping Motifs

No.	Pos.	Group	Score	No.	Pos.	Group	Score
1	K254	PVCGD **IKVE** FFHKQ	0.94	4	K289	PEETS **EKVE** NGSLC	0.5
2	K266	KQNKM **LKKD** KMFHF	0.91	5	K163	EVRTR **DKKG** VTIPS	0.33
3	K102	PQLEL **IKPF** CEDLD	0.59	6	K269	KMLKK **DKMF** HFWVN	0.15

Experiment evidence suggested that lysine residue 266 and 289 as SUMO acceptors .

González-Santamaría J et al. ***Cell Death Dis***. 2012 Sep 27;3:e393.

Summary

- Swiss-Prot/ Uniprot
- Protein sequences and identification
- Protein structure prediction
- Protein function
- Experiment

Acknowledgements

- Prof. Luo.
- Dr. Wang.
- Cai Tiantian, Liu Yiqiong, Jiang Qianru.